IJHSR International Journal of High School Research

August 2022 | Volume 4 | Issue 4
ijhighschoolresearch.org
ISSN (Print) 2642-1046
ISSN (Online) 2642-1054

GENIUS OLYMPIAD

"Let's build a better future together"

www.geniusolympiad.org

International Environment Project Fair For Grades 9-12

Rochester, New York

Hosted by Rochester Institute of Technology

RIT Rochester Institute of Technology

@GeniusOlympiad

terra science and education

Table of Contents

August 2022| Volume 4| Issue 4

Editorial Board

International Journal of High School Research

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

A Study of Gene Expression in the Brains of COVID-19, Alzheimer's Disease, and Vascular Dementia Patients

Aden Geonhee Lee
Phillips Exeter Academy, 20 Main St, Exeter, NH 03833,USA; galee@exeter.edu

ABSTRACT: COVID-19 is associated with neurological complications, like the loss of smell, stroke, and dementia. However, it is not well understood how COVID-19 infects brain cells and which processes it affects. The purpose of this study was to investigate the differential and overlapping gene expression in COVID-19, Alzheimer's disease (AD), and Vascular Dementia (VD). Two gene expression datasets were collected from Gene Expression Omnibus. These datasets contained postmortem frontal cortex samples from COVID-19, AD, VD patients, and controls. Then, differential expressed genes (DEGs) were analyzed using GEO2R and Galaxy. Finally, enriched function and pathways were analyzed by DAVID, protein-protein interaction by STRING, and protein-protein docking by HADDOCK. C-C motif chemokine ligand 2 (CCL2), overlapped in three diseases, and was most enriched in the inflammatory pathway. In the docking analysis, the NMR structure of the nonstructural protein 1 (NSP1) from COVID-19 strongly bound to human ribosomal protein S11 (RPS11). These findings suggest that the NSP1 of COVID-19 may directly bind to the brain RPS11, and indirectly cause an inflammatory reaction through CCL2 in the brain.

KEYWORDS: Computational Biology and Bioinformatics; Genomics; COVID-19 neuropathology; Alzheimer's disease; Vascular dementia.

Introduction

COVID-19, a disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection,[1] is primarily characterized by respiratory symptoms. The most common symptoms are fever, cough, fatigue, and shortness of breath, but COVID-19 patients also can suffer neurological and psychiatric symptoms.[2,3] COVID-19 neurological symptoms can include seizures, cerebrovascular accidents, encephalopathy, and Guillain-Barré Syndrome.[4] Many of the neurological complications of SARS-CoV-2 infection are related to severe systemic conditions, including multiple organ damage, hypercoagulability, over-activated inflammatory response, and direct infection of the brain.[5]

Multiple studies have found evidence of SARS-CoV-2 in brain samples of COVID-19 infected patients,[6-8] suggesting that SARS-CoV-2 may be responsible for these neurological complications. Indirectly, COVID-19 neuropathology has also been hypothesized to arise as a result of moderate to severe hypoxemia, metabolic dysfunction, systemic inflammation, and immune dysregulation that can contribute to brain dysfunction, all of which have been documented in patients.[9,10] Alzheimer's disease and Vascular dementia, which are common causes of brain degeneration, have been shown to increase the mortality and hospitalization risk for COVID-19 patients.[11] Further, observed activation of microglia in the brains of COVID-19 patients can also be found in patients with neurodegenerative diseases, including Vascular dementia and Alzheimer's disease.[12] These studies suggest that COVID-19 may have unique neuropathology, or an overlapping pathology with other central nervous system diseases.

To date, there has been little research on the effects SARS-CoV-2 on the brain, nor the cellular and molecular mechanisms underlying the neurological changes observed in COVID-19 patients. This study investigated the differential gene expression in postmortem brains of COVID-19 patients and compared them with those of Alzheimer's disease and Vascular dementia patients to identify genes and pathways that may be shared between these three diseases. The common pathways and genes revealed in this study will hopefully provide insight into the neuropathology of COVID-19 and provide candidate targets for future neuro-protective studies in COVID-19 patients.

Methods

Data Sources:

Two datasets that were selected in this study were collected from NCBI's Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/). The first dataset, GSE164332, was created from a study that analyzed post-mortem brain samples of COVID-19, and age and clinical characteristics-matched non-COVID-19 control subjects.[13] Although the post-mortem time of COVID-19 was relatively longer, all the bodies were adequately preserved until the time of the autopsy. The second dataset, GSE122063, was from the study including post-mortem brain samples of Alzheimer's disease, Vascular Dementia and age-matched controls.[14] The GSE164332 dataset investigated the total RNA expression from postmortem brain frontal lobe from nine COVID-19 patients and eight non-COVID-19 controls for comparison, which included patients with Alzheimer's disease and Vascular Dementia. Therefore, only six cases (COVID-19 = 3, non-COVID-19 = 3) without Alzheimer's disease, Vascular Dementia (Table 1) were selected. Microarray analysis was performed using Illumina NextSeq 500 Sequencing, and FastQ files were generated via Illumina bcl2fastq2 (Version 2.17.1.14). Gene expression levels were also quantified and normalized. The GSE122063 dataset analyzed the total

DOI: 10.36838/v4i4.1

normalized. The GSE122063 dataset analyzed the total RNA extracted from postmortem brain frontal lobe from individuals who died with Alzheimer's disease (n = 12), Vascular Dementia (primarily of the multi-infarct dementia subtype, n = 9), or without dementia controls (n = 10) (Table 1). Alzheimer's disease was confirmed by no infarct lesions in the autopsied brain hemisphere, and Vascular Dementia was confirmed by no evidence of Alzheimer's disease typical pathological feature. The age, sex, and postmortem interval hours-matched controls were selected. Agilent Human 8x60k v2 microarrays and Bioconductor were used to perform gene expression analysis, and differential expression was carried out using limma.

Table 1: Characteristics of study population from GEO and GSE datasets.

Groups	Age	Sex (M:F)	Postmortem interval (hours)	GEO dataset
COVID-19 (n=3)	79 ± 9.4	1:2	248 ± 98.6	GSE164332
Non-COVID-19 control (n=3)	80.6 ± 2.4	2:1	9.6 ± 5.3	GSE164332
Alzheimer's disease (n=12)	80.9 ± 7.4	4:5	8.0 ± 4.0	GSE122063
Vascular Dementia (n=9)	81.4 ± 10.1	5:7	10.0 ± 4.0	GSE122063
Control (n=10)	78.6 ± 8.5	5:5	9.0 ± 3.0	GSE122063

Differential Gene Analysis:

Galaxy (https://usegalaxy.org/), an online tool that analyzes the gene expression between groups in a GEO dataset, was used to determine the differential expressed genes in GSE164332, and GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r/) in GSE122063. Analyzed results from Galaxy and GEO2R for each dataset were saved to text format files. Genes with a false discovery rate (FDR)-adjusted $p < 0.05$, and a $\log_2$ fold change (FC) cutoff of 1.0 were considered differentially expressed. Genes that met both FDR and logFC cutoffs for each dataset were transported to Excel and used in enriched function analysis. The overlapping genes in the three disease groups were identified using Venny 2.1 (https://bioinfogp.cnb.csic.es/tools/venny/).

Enriched Function Analysis with DAVID and STRING:

The functions and pathway enrichment of candidate DEGs were analyzed using the Database for Annotation, Visualization and Integrated Discovery database (DAVID, https://david.ncifcrf.gov/), which is a web-accessible program that accommodates a comprehensive set of functional annotation tools to investigate the biological roles of genes. The results are presented as dot plots using R Studio and ggplot2 with the enriched functional terms of the molecular function (MF), cellular components (CC), biological process (BP) in Gene ontology (GO) analysis, and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. The identified genes from STRING were inputted into DAVID, and $p < 0.05$ was considered as significant. In addition, a count≥2 and EASE >0.1 were considered the cut-off criteria. To analyze the enriched functions that each gene belongs to, the STRING (https://string-db.org/), which predicts direct and indirect protein interactions using online databases and presents a diagram showing important protein–protein interactions (PPI), were used. The differential expressed genes from differential gene analysis were inputted into the STRING, and the minimum interaction threshold was set to the "highest confidence" (>0.9).

Protein-Protein Docking Analysis:

The interactions between SARS-CoV-2 and their target proteins were analyzed to evaluate their binding site and affinity. RPS11 (PDB: 6ZLW) of human 40s ribosome and NSP1 (PDB: 2HSX) of SARS-CoV-2, which was obtained from Protein Data Bank (https://www.rcsb.org/) were selected. These selected proteins were inputted in High Ambiguity Driven protein-protein Docking (HADDOCK, https://wenmr.science.uu.nl/haddock2.4/), a data-driven flexible docking program for the modeling of protein-protein biomolecular complexes. The result of protein-protein Docking was presented by RMSD.

Results

Differential gene expression:

To examine the differentially expressed genes in the postmortem brain of COVID-19 patients as compared to those of non-COVID-19 controls, Galaxy analysis of GSE164332 was performed.[9] 79 over-expressed and 472 down-expressed genes were identified in COVID-19 brain tissue compared to non-COVID-19 controls. The gene *KIAA0319*, which is associated with dyslexia[15] and encodes a protein that plays a role in the developing cerebral cortex by neuronal migration and cell adhesion,[16] was the most over-expressed gene in COVID-19 patients (Table 2). *Myoferlin* (*MYOF*), which encodes a protein that plays a role in membrane regeneration and repair by calcium-mediated membrane fusion,[17] was the most down-expressed gene in COVID-19 patient (Table 3).

Next, GEO2R analysis of GSE122063 was conducted. GSE122063 expression data was derived from a study that compared postmortem frontal lobe samples from Alzheimer's disease or Vascular Dementia patients with samples from controls.[14] This second analysis returned 526 over-expressed and 408 down-expressed genes between Alzheimer's disease patients and control individuals, and 193 over-expressed and 170 down-expressed genes between Vascular Dementia patients and control individuals. In the brains of Alzheimer's disease patients, *corticotropin releasing hormone* (*CRH*) was the most over-expressed gene (Table 2), and *interleukin 1 receptor like 1* (*IL1RL1*), induced by proinflammatory stimuli,[18] was the most down-expressed (Table 3). In the brains of Vascular Dementia patients, *RNA binding motif protein 3* (*RBM3*) gene, induced by low oxygen tension,[19] had the highest expression (Table 2), and *sorting nexin 31* (*SNX31*) gene, involved in protein trafficking,[20] had the lowest expression (Table 3).

Overlapping genes:

In order to determine if there were common gene pathways involved in COVID-19, Alzheimer's disease, and Vascular Dementia patients, genes were screened with conditions of the FDR-adjusted p-value cutoff of 0.05 and logFC cutoff of 1.0 in each dataset, and candidates that overlapped across

Table 2: The top 10 over-expressed genes in postmortem brains of COVID-19, Alzheimer's disease, and Vascular Dementia patients.

COVID-19		Alzheimer dis		Vascular Dementia	
Gene	*logFC*	*Gene*	*logFC*	*Gene*	*logFC*
KIAA0319	3.013166	CRH	2.95	RBM3	2.509663
CDC14C	2.458743	CARTPT	2.81	STMN1	2.150447
ZACN	2.203234	FRMPD2	2.59	CCT6B	2.014873
NPIPA5	2.113072	VGF	2.37	RBM3	1.999349
WASH5P	1.89861	SST	2.32	PBOV1	1.980297
YWHAEP5	1.828757	CTXN3	2.28	LINC00458	1.841953
TP73-AS1	1.806999	GAD2	2.26	PI3	1.744062
DDX18P5	1.754231	PPEF1	2.23	CRH	1.735625
PIM3	1.702827	RBM3	2.13	GUCY2GP	1.720652
PRXL2B	1.680065	PCSK1	2.09	KLHDC7B	1.687498

Table 3: The top 10 down-expressed genes in postmortem brains of COVID-19, Alzheimer's disease and Vascular Dementia patients.

COVID-19		Alzheimer dis		Vascular Dementia	
Gene	*logFC*	*Gene*	*logFC*	*Gene*	*logFC*
MYOF	-5.2812	IL1RL1	-2.7	SNX31	-1.91756
GEMIN5	-5.00056	SERPINA3	-2.48	FCGBP	-1.91508
HDGFL2	-4.95203	SERPINA3	-2.47	SLAMF8	-1.90554
LTA4H	-4.86781	AQP1	-2.29	MIA	-1.88122
GNPAT	-4.45946	SLC1A7	-2.12	CD163	-1.79706
AC104457.2	-4.39108	FCGBP	-2.08	VSIG4	-1.7718
FAN1	-4.36097	CCL2	-2	SIGLEC14	-1.74045
CEP135	-4.32272	SLAMF8	-1.91	SPP1	-1.68427
SERPINI2	-4.27884	HMBOX1	-1.9	RNASE2	-1.61027
KMO	-4.25159	CHRM4	-1.88	AQP1	-1.6017

COVID-19, Alzheimer's disease, and Vascular Dementia were collected. This analysis revealed 67 differentially over-expressed genes and 105 differentially down-expressed genes that overlap in COVID-19, Alzheimer's disease, and/or Vascular Dementia patients (Figure 1). In particular, three over-expressed genes in Alzheimer's disease samples that overlapped with COVID-19 that were of interest included *neuronal pentraxin 2* (*NPTX2*), growth-regulating estrogen receptor binding 1 (*GREB1*) and *allograft inflammatory factor 1-like* (*AIF1L*) (Figure 1A). Additionally, for COVID-19 and Alzheimer's disease, five down-expressed overlapping genes were found: *posphatidate phosphatase LPIN30* (*LPIN3*), *serglycin* (*SRGN*), *solute carrier family 7 member 2* (*SLC7A2*), *caspase recruitment domain family member 6* (*CARD6*), and *Ral guanine nucleotide dissociation stimulator like 4* (*RGL4*) (Figure 1B). In all three diseases, *ribonuclease A family member 2* (*RNASE2*) and *C-C motif chemokine ligand 2* (*CCL2*) were identified (Figure 1B). These results suggest that inflammation-related genes are involved in the pathology of COVID-19, Alzheimer's disease, and Vascular dementia.

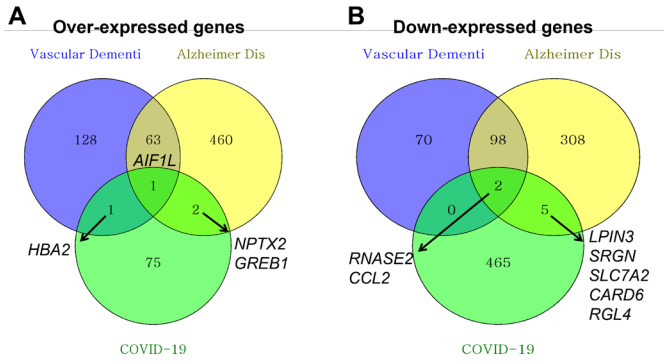


Figure 1: Number of overlapping genes of (A) over-expressed genes and (B) down-expressed genes in brain samples from COVID-19 (green), Alzheimer's disease (yellow), and Vascular Dementia (blue) patients.

Statistical Evaluation:

To analyze the enriched functions and pathways that differential expressed and overlapping genes belong to, 551 differentially expressed genes of COVID-19, 934 genes of Alzheimer's disease, 363 genes of Vascular Dementia, and 182 overlapping genes of the three diseases were analyzed using enriched gene ontology (GO) analysis. The enriched functions and pathways of the candidate differentially expressed genes were evaluated using the DAVID bioinformatics website.

Analysis revealed differentially expressed genes (DEGs) of COVID-19 samples were enriched in viral transcription, viral gene expression, and co-translational protein targeting to membrane pathways in the BP ontology (Figure 2A). In the CC ontology, aberrantly expressed genes were enriched in the ribosomal subunit and cytosolic ribosome. In the MF ontology, structural constituents of ribosomes were the main areas of enrichment. According to the KEGG pathway enrichment analysis, DEGs were enriched in transcriptional misregulation and the ribosome signaling pathway in COVID-19 samples (Figure 2A). These data indicate that the coronavirus infection is associated with changes in gene expression and specifically the ribosomal pathway.

In the Alzheimer's disease samples, the pathways of signaling, cell communication, nervous system development, and synaptic signaling were enriched in the BP ontology while the pathways involved in synapse and neuron projection were enriched in the CC ontology (Figure 2B). In addition, neurotransmitter receptor activity was enriched in the MF ontology (Figure 2B). Lastly, the pathways involved in neuroactive ligand-receptor interactions as well as inhibitory GABAergic synapses were enriched in KEGG pathway (Figure 2B). These data indicate that genes involved in excitatory and inhibitory synaptic neurotransmission are upregulated in the brains of Alzheimer's patients. In Vascular Dementia, G protein-coupled receptor signaling pathways were enriched in the BP ontology, and Neuroactive ligand-receptor interaction were enriched in KEGG pathway (Figure 2C).

The overlapping pathways enriched in COVID-19, Alzheimer's disease, and Vascular Dementia included chemotaxis, immune response, inflammatory response, positive regulation of immune complex clearance by monocytes and macrophages, regulation of isotype switching, regulation of vascular endothelial growth factor production, type 2 immune response, and positive regulation of the major histocompatibility complex (MHC) class II biosynthetic process in the BP ontology (Figure 2D). Gene products within the extracellular region and extracellular space were enriched in the CC ontology, while CCR2 chemokine receptor binding and ribonuclease activity were enriched in the MF ontology, and cytokine-cytokine receptor interaction were enriched in KEGG pathway (Figure 2D). These data indicate that inflammation, particularly related with monocyte/macrophages and CCR2, is a common pathological feature in these three diseases.

Protein-Protein Interaction:

To construct a protein-protein interaction (PPI) network of differential expressed and overlapping genes, the enriched PPI were analyzed. 551 differentially expressed genes of

DOI: 10.36838/v4i4.1

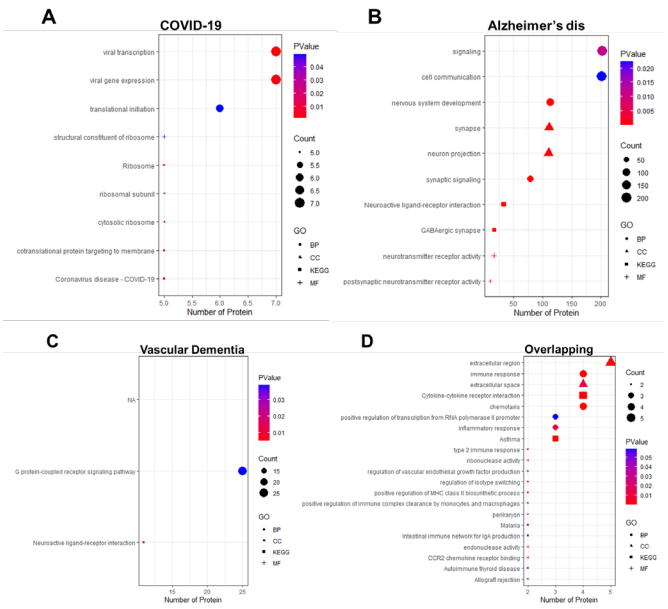


Figure 2: Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway of (A) COVID-19, (B) Alzheimer's disease, (C) Vascular dementia, and (D) the overlapping genes enriched in these three diseases.

COVID-19, 934 genes of Alzheimer's disease, 363 genes of Vascular Dementia, and 182 common genes of three diseases were inputted into the STRING to find PPI. PPI analysis revealed that *ribosomal protein S11* (*RPS11*) of COVID-19, *brain-derived neurotrophic factor* (*BDNF*) of Alzheimer's disease, *G protein subunit gamma 13* (*GNG13*) of Vascular Dementia, and *C-C motif chemokine ligand 2* (*CCL2*) were identified as highest-ranked interacting hub genes (Figure 3).

The *RPS11* hub gene in COVID-19 samples was chiefly enriched in the ribosome pathway, and *BDNF* in Alzheimer's disease samples was chiefly enriched in the GABAergic synapse pathway, while *GNG13* in Vascular Dementia was chiefly enriched in G protein-coupled receptor signaling pathway (Figure 3A-3C). The common hub gene of these three diseases, *CCL2*, was chiefly enriched in cytokine-cytokine receptor interaction (Figure 3D).

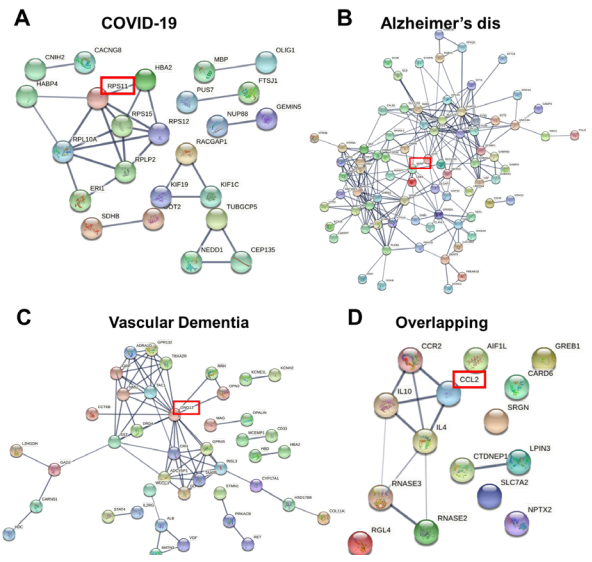


Figure 3: Protein-Protein Interactions of (A) COVID-19, (B) Alzheimer's disease, (C) Vascular dementia, and (D) overlapping genes of these three diseases.

Protein-Protein docking:

To evaluate whether the RPS11 hub protein from enriched functional analysis could interact with the SARS CoV-2 protein, molecular docking tools were employed to analyze the interaction of the Nonstructural protein 1 (NSP1) from the SARS coronavirus (PDB: 2HSX) with human RPS11 (PDB: 6ZLW). In this study, the SARS-CoV N-terminal NSP1 was selected because it is a major virulence factor of SARS-CoV-2.[21] Protein-protein interaction complexes were analyzed using the HADDOCK 2.4 server which is a computational tool that allows the simulation of protein-protein molecular interaction. As a result, the HADDOCK energy score of this interaction was -75.9 ± 1.7, overall RMSD was 0.5 ± 0.4Å, Van der Waals energy was -23.9 ± 1.7, and the Z-score was -1.7 (Figure 4).

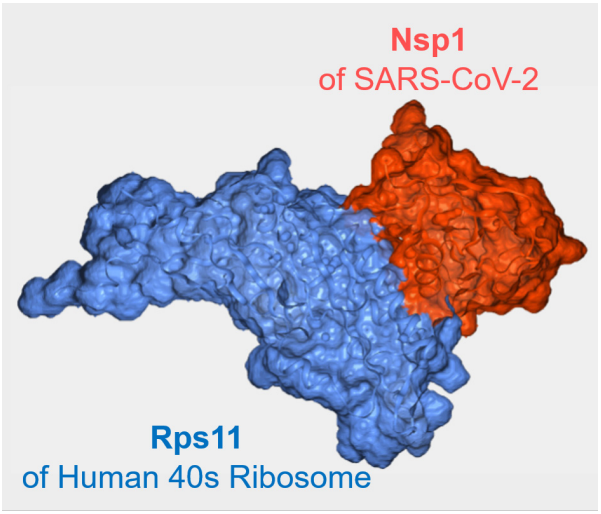


Figure 4: Protein-Protein docking of NMR Structure of the nonstructural protein 1 (NSP1) from the SARS coronavirus (PDB: 2HSX) and human ribosomal protein S11 (RPS11) (PDB: 6ZLW).

■ Discussion

The data presented here shows that in the brains of COVID-19, Alzheimer's disease, and Vascular dementia patients the genes *RPS11* of COVID-19, *BDNF of Alzheimer's disease*, *GNG13* of Vascular Dementia, and *CCL2* of all three diseases are interacting hub genes when compared to control individuals. To hypothesize the role of these genes in a disease context requires some background into what is known about these genes.

RPS11, enriched in COVID-19 samples, encodes a ribosomal protein S11 located in the cytoplasm.[22] Here, it can be seen that there is a strong interaction between the RPS11 protein and the SARS CoV-2 protein using molecular docking tools. Indeed, another group has reported that NSP1, a major virulence factor of SARS-CoV-2, binds to the 40s ribosomal subunit of RPS11 and suppresses host gene expression.[21] Given that other studies have also detected SARS-CoV-2 in brain tissue,[6, 7, 23] this presents an intriguing potential mechanism for COVID-19 neuropathology. Specifically, if NSP1 of SARS-CoV-2 could invade the human brain and bind to 40s

ribosomal subunit of RPS11, this could result in impaired ribosome function, and consequently, decreased protein synthesis in neurons leading to neurodegeneration, and potentially the brain fog that is experienced by patients.

Here it was found that *BDNF* is enriched in the GABAergic synapses in Alzheimer's disease brain samples. The *BDNF* gene encodes a protein which modulates control of growth, differentiation, and maintenance of neurons in the brain and spinal cord.[24] BDNF regulates the synaptic plasticity between nerve cells, playing an important role in learning and memory.[25] In Alzheimer's disease, amyloid-beta is thought to target GABA interneurons[26] which may result in indirect activation of BDNF[27] and a shift in the balance of excitatory/inhibitory activity in the brains of patients. In this study, *GNG13* was found to be chiefly enriched in the G protein-coupled receptor signaling pathway in Vascular Dementia samples. *GNG13* encodes the guanine nucleotide-binding protein subunit gamma-13. G proteins, like GNG13, are signal transducing molecules in the cytoplasm. G proteins can function as monomeric small GTPases or form heterotrimeric G protein complexes (GPCRs) to function as membrane bound GTPases. G proteins and GPCRs transfer the signals from hormones, neurotransmitters, and chemokines, as well as modulate enzymes, neurochemistry, and transcriptional factors.[28] The disruption of GPCRs are implicated in many diseases, including diabetes, allergies, depression, cardiovascular diseases,[29] and neurodegenerative diseases such as Vascular Dementia.[30] Vascular Dementia is caused by restricted blood flow to the brain, which can be caused by impairment of the cerebral blood vessel system, resulting in neuron death.[31] GPCRs play important roles in various pathophysiological disorders including circulatory diseases[28] and not surprisingly, are commonly targeted for medicinal therapeutics.

Interestingly, the CCL2 inflammatory gene and inflammatory response pathway were enriched in COVID-19, Alzheimer's disease, and Vascular dementia. CCL2, or Monocyte chemoattractant protein 1 (MCP1), is a chemokine which recruits inflammatory monocytes to the inflammation site induced by tissue injury or infection.[32] There is increasing evidence showing overlap of cerebrovascular and neurodegenerative pathology in Alzheimer's disease and Vascular Dementia.[33] In Alzheimer's disease, CCL2 is increased both in brain tissue and in cerebrospinal fluid (CSF), and overexpression of CCL2 is associated with amyloid deposition.[34] In the ischemic state of Vascular Dementia, CCL2 disrupts blood-brain barrier (BBB) permeability, recruits inflammatory monocytes to the brain, where they pass through BBB and are differentiated into microglia producing neurotoxic and inflammatory molecules.[35, 36] The systemic cytokine analysis of severe COVID-19 showed the significantly higher level of CCL2.[37] Furthermore, an elevated levels of CCL2 in the CSF of COVID-19 patients were reported.[38]

SARS-CoV-2 in the brain results in a markable increase of microglia and astrocyte subpopulations and broad over-expression of inflammatory genes in the choroid plexus cells of COVID-19 patients.[11] This microglia phenotype shares features with neurodegenerative diseases that have previously been reported. These results, in combination with the findings of this study, suggest that SARS-CoV-2 infection could directly affect the brain, cause inflammation, and possibly cause a patient's neurological symptoms.

Conclusion

Taken together, this research has provided evidence linking important pathways that may function in the pathology of COVID-19, Alzheimer's disease, and Vascular dementia. From this data, it is hypothesized that, NSP1 of SARS-CoV-2 binds to RPS of 40s ribosome in neurons, suppressing gene expression, and contributing to CCL2 inflammatory reaction in the cerebral cortex, resulting in neuronal degeneration. To confirm this hypothesis, further studies will need to be conducted to support these links, but the data gathered here will hopefully be utilized for future studies into therapeutics for the preservation of brain function in COVID-19 patients.

Acknowledgements

Aden Geonhee Lee thanks Heather Killeen, Ph.D. for helping with revisions to the manuscript prior to publication.

References

1. Zou, L.; Ruan, F.; Huang, M.; Liang, L.; Huang, H.; Hong, Z.; Yu, J.; Kang, M.; Song, Y.; Xia, J.; Guo, Q.; Song, T.; He, J.; Yen, H. L.; Peiris, M.; Wu, J., SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. *N Engl J Med* **2020**, 382 (12), 1177-1179.
2. Helms, J.; Kremer, S.; Merdji, H.; Clere-Jehl, R.; Schenck, M.; Kummerlen, C.; Collange, O.; Boulay, C.; Fafi-Kremer, S.; Ohana, M.; Anheim, M.; Meziani, F., Neurologic Features in Severe SARS-CoV-2 Infection. *N Engl J Med* **2020,** 382 (23), 2268-2270.
3. Marshall, M., How COVID-19 can damage the brain. *Nature* **2020**, 585 (7825), 342-343.
4. Iadecola, C.; Anrather, J.; Kamel, H., Effects of COVID-19 on the Nervous System. *Cell* **2020**, *183* (1), 16-27 e1.
5. Thakur, K. T.; Miller, E. H.; Glendinning, M. D.; Al-Dalahmah, O.; Banu, M. A.; Boehme, A. K.; Boubour, A. L.; Bruce, S. S.; Chong, A. M.; Claassen, J.; Faust, P. L.; Hargus, G.; Hickman, R. A.; Jambawalikar, S.; Khandji, A. G.; Kim, C. Y.; Klein, R. S.; Lignelli-Dipple, A.; Lin, C. C.; Liu, Y.; Miller, M. L.; Moonis, G.; Nordvig, A. S.; Overdevest, J. B.; Prust, M. L.; Przedborski, S.; Roth, W. H.; Soung, A.; Tanji, K.; Teich, A. F.; Agalliu, D.; Uhlemann, A. C.; Goldman, J. E.; Canoll, P., COVID-19 neuropathology at Columbia University Irving Medical Center/New York Presbyterian Hospital. *Brain* **2021**.
6. Matschke, J.; Lutgehetmann, M.; Hagel, C.; Sperhake, J. P.; Schroder, A. S.; Edler, C.; Mushumba, H.; Fitzek, A.; Allweiss, L.; Dandri, M.; Dottermusch, M.; Heinemann, A.; Pfefferle, S.; Schwabenland, M.; Sumner Magruder, D.; Bonn, S.; Prinz, M.; Gerloff, C.; Puschel, K.; Krasemann, S.; Aepfelbacher, M.; Glatzel, M., Neuropathology of patients with COVID-19 in Germany: a post-mortem case series. *Lancet Neurol* **2020**, *19* (11), 919-929.
7. Song, E.; Zhang, C.; Israelow, B.; Lu-Culligan, A.; Prado, A. V.; Skriabine, S.; Lu, P.; Weizman, O. E.; Liu, F.; Dai, Y.; Szigeti-Buck, K.; Yasumoto, Y.; Wang, G.; Castaldi, C.; Heltke, J.; Ng, E.; Wheeler, J.; Alfajaro, M. M.; Levavasseur, E.; Fontes, B.; Ravindra, N. G.; Van Dijk, D.; Mane, S.; Gunel, M.; Ring, A.; Kazmi, S. A. J.; Zhang, K.; Wilen, C. B.; Horvath, T. L.; Plu, I.; Haik, S.; Thomas, J. L.; Louvi, A.; Farhadian, S. F.; Huttner, A.; Seilhean, D.; Renier, N.; Bilguvar, K.; Iwasaki, A., Neuroinvasion of SARS-CoV-2 in human and mouse brain. *J Exp Med* **2021**, *218* (3).
8. Pellegrini, L.; Albecka, A.; Mallery, D. L.; Kellner, M. J.; Paul,

DOI: 10.36838/v4i4.1

D.; Carter, A. P.; James, L. C.; Lancaster, M. A., SARS-CoV-2 Infects the Brain Choroid Plexus and Disrupts the Blood-CSF Barrier in Human Brain Organoids. *Cell Stem Cell* **2020**, *27* (6), 951-961 e5.
9. Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; Cheng, Z.; Yu, T.; Xia, J.; Wei, Y.; Wu, W.; Xie, X.; Yin, W.; Li, H.; Liu, M.; Xiao, Y.; Gao, H.; Guo, L.; Xie, J.; Wang, G.; Jiang, R.; Gao, Z.; Jin, Q.; Wang, J.; Cao, B., Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **2020**, *395* (10223), 497-506.
10. Guan, W. J.; Ni, Z. Y.; Hu, Y.; Liang, W. H.; Ou, C. Q.; He, J. X.; Liu, L.; Shan, H.; Lei, C. L.; Hui, D. S. C.; Du, B.; Li, L. J.; Zeng, G.; Yuen, K. Y.; Chen, R. C.; Tang, C. L.; Wang, T.; Chen, P. Y.; Xiang, J.; Li, S. Y.; Wang, J. L.; Liang, Z. J.; Peng, Y. X.; Wei, L.; Liu, Y.; Hu, Y. H.; Peng, P.; Wang, J. M.; Liu, J. Y.; Chen, Z.; Li, G.; Zheng, Z. J.; Qiu, S. Q.; Luo, J.; Ye, C. J.; Zhu, S. Y.; Zhong, N. S.; China Medical Treatment Expert Group for, C., Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* **2020**, *382* (18), 1708-1720.
11. Wang, Q.; Davis, P. B.; Gurney, M. E.; Xu, R., COVID-19 and dementia: Analyses of risk, disparity, and outcomes from electron ic health records in the US. *Alzheimers Dement* **2021**, *17* (8), 1297-1306.
12. Yang, A. C.; Kern, F.; Losada, P. M.; Agam, M. R.; Maat, C. A.; Schmartz, G. P.; Fehlmann, T.; Stein, J. A.; Schaum, N.; Lee, D. P.; Calcuttawala, K.; Vest, R. T.; Berdnik, D.; Lu, N.; Hahn, O.; Gate, D.; McNerney, M. W.; Channappa, D.; Cobos, I.; Ludwig, N.; Schulz-Schaeffer, W. J.; Keller, A.; Wyss-Coray, T., Dysregulation of brain and choroid plexus cell types in severe COVID-19. *Nature* **2021**, *595* (7868), 565-571.
13. Gagliardi, S.; Poloni, E. T.; Pandini, C.; Garofalo, M.; Dragoni, F.; Medici, V.; Davin, A.; Visona, S. D.; Moretti, M.; Sprovie ro, D.; Pansarasa, O.; Guaita, A.; Ceroni, M.; Tronconi, L.; Cereda, C., Detection of SARS-CoV-2 genome and whole tran scriptome sequencing in frontal cortex of COVID-19 patients. *Brain Behav Immun* **2021**.
14. McKay, E. C.; Beck, J. S.; Khoo, S. K.; Dykema, K. J.; Cotting ham, S. L.; Winn, M. E.; Paulson, H. L.; Lieberman, A. P.; Counts, S. E., Peri-Infarct Upregulation of the Oxytocin Receptor in Vascular Dementia. *J Neuropathol Exp Neurol* **2019**, *78* (5), 436-452.
15. Ludwig, K. U.; Roeske, D.; Schumacher, J.; Schulte-Korne, G.; Konig, I. R.; Warnke, A.; Plume, E.; Ziegler, A.; Remschmidt, H.; Muller-Myhsok, B.; Nothen, M. M.; Hoffmann, P., Investiga tion of interaction between DCDC2 and KIAA0319 in a large German dyslexia sample. *J Neural Transm (Vienna)* **2008**, *115* (11), 1587-9.
16. Velayos-Baeza, A.; Levecque, C.; Kobayashi, K.; Holloway, Z. G.; Monaco, A. P., The dyslexia-associated KIAA0319 protein undergoes proteolytic processing with {gamma}-secretase-inde pendent intramembrane cleavage. *J Biol Chem* **2010**, *285* (51), 40148-62.
17. Dong, Y.; Kang, H.; Liu, H.; Wang, J.; Guo, Q.; Song, C.; Sun, Y.; Zhang, Y.; Zhang, H.; Zhang, Z.; Guan, H.; Fang, Z.; Li, F., Myoferlin, a Membrane Protein with Emerging Oncogenic Roles. *Biomed Res Int* **2019**, *2019*, 7365913.
18. Tajima, S.; Oshikawa, K.; Tominaga, S.; Sugiyama, Y., The in crease in serum soluble ST2 protein upon acute exacerbation of id iopathic pulmonary fibrosis. *Chest* **2003**, *124* (4), 1206-14.
19. Wellmann, S.; Buhrer, C.; Moderegger, E.; Zelmer, A.; Kirschner, R.; Koehne, P.; Fujita, J.; Seeger, K., Oxygen-regulated expression of the RNA-binding proteins RBM3 and CIRP by a HIF-1-independent mechanism. *J Cell Sci* **2004**, *117* (Pt 9), 1785-94.
20. Vieira, N.; Deng, F. M.; Liang, F. X.; Liao, Y.; Chang, J.; Zhou, G.; Zheng, W.; Simon, J. P.; Ding, M.; Wu, X. R.; Romih, R.; Kreibich, G.; Sun, T. T., SNX31: a novel sorting nexin associated with the uroplakin-degrading multivesicular bodies in terminally differentiated urothelial cells. *PLoS One* **2014**, *9* (6), e99644.
21. Thoms, M.; Buschauer, R.; Ameismeier, M.; Koepke, L.; Denk, T.; Hirschenberger, M.; Kratzat, H.; Hayn, M.; Mackens-Ki ani, T.; Cheng, J.; Straub, J. H.; Sturzel, C. M.; Frohlich, T.; Berninghausen, O.; Becker, T.; Kirchhoff, F.; Sparrer, K. M. J.; Beckmann, R., Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science* **2020**, *369* (6508), 1249-1255.
22. Kenmochi, N.; Kawaguchi, T.; Rozen, S.; Davis, E.; Goodman, N.; Hudson, T. J.; Tanaka, T.; Page, D. C., A map of 75 human ribosomal protein genes. *Genome Res* **1998**, *8* (5), 509-23.
23. Puelles, V. G.; Lutgehetmann, M.; Lindenmeyer, M. T.; Sper hake, J. P.; Wong, M. N.; Allweiss, L.; Chilla, S.; Heinemann, A.; Wanner, N.; Liu, S.; Braun, F.; Lu, S.; Pfefferle, S.; Schro der, A. S.; Edler, C.; Gross, O.; Glatzel, M.; Wichmann, D.; Wiech, T.; Kluge, S.; Pueschel, K.; Aepfelbacher, M.; Huber, T. B., Multiorgan and Renal Tropism of SARS-CoV-2. *N Engl J Med* **2020**, *383* (6), 590-592.
24. Binder, D. K.; Scharfman, H. E., Brain-derived neurotrophic fac tor. *Growth Factors* **2004,** *22* (3), 123-31.
25. Rodriguez-Lopez, R.; Perez, J. M.; Balsera, A. M.; Rodriguez, G. G.; Moreno, T. H.; Garcia de Caceres, M.; Serrano, M. G.; Freijo, F. C.; Ruiz, J. R.; Angueira, F. B.; Perez, P. M.; Estevez, M. N.; Gomez, E. G., The modifier effect of the BDNF gene in the phenotype of the WAGRO syndrome. *Gene* **2013**, *516* (2), 285-90.
26. Villette, V.; Dutar, P., GABAergic Microcircuits in Alzheimer's Disease Models. *Curr Alzheimer Res* **2017**, *14* (1), 30-39.
27. Giuffrida, M. L.; Copani, A.; Rizzarelli, E., A promising con nection between BDNF and Alzheimer's disease. *Aging (Albany NY)* **2018**, *10* (8), 1791-1792.
28. Neves, S. R.; Ram, P. T.; Iyengar, R., G protein pathways. *Science* **2002**, *296* (5573), 1636-9.
29. Bosch, D. E.; Siderovski, D. P., G protein signaling in the parasite Entamoeba histolytica. *Exp Mol Med* **2013**, *45*, e15.
30. Huang, Y.; Todd, N.; Thathiah, A., The role of GPCRs in neuro degenerative diseases: avenues for therapeutic intervention. *Curr Opin Pharmacol* **2017**, *32*, 96-110.
31. Karczewski, P.; Hempel, P.; Bimmler, M., Role of alpha1-adren ergic receptor antibodies in Alzheimer's disease. *Front Biosci (Landmark Ed)* **2018**, *23*, 2082-2089.
32. Deshmane, S. L.; Kremlev, S.; Amini, S.; Sawaya, B. E., Monocyte chemoattractant protein-1 (MCP-1): an overview. *J Interferon Cy tokine Res* **2009**, *29* (6), 313-26.
33. de la Torre, J. C., Alzheimer disease as a vascular disorder: nosolog ical evidence. *Stroke* **2002**, *33* (4), 1152-62.
34. Azizi, G.; Khannazer, N.; Mirshafiey, A., The Potential Role of Chemokines in Alzheimer's Disease Pathogenesis. *Am J Alzhei mers Dis Other Demen* **2014**, *29* (5), 415-25.
35. Iadecola, C., The pathobiology of vascular dementia. *Neuron* **2013**, *80* (4), 844-66.
36. Dimitrijevic, O. B.; Stamatovic, S. M.; Keep, R. F.; Andjelkovic, A. V., Effects of the chemokine CCL2 on blood-brain barrier per meability during ischemia-reperfusion injury. *J Cereb Blood Flow Metab* **2006**, *26* (6), 797-810.
37. Merad, M.; Martin, J. C., Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. *Nat Rev Immunol* **2020,** *20* (6), 355-362.
38. Savarraj, J.; Park, E. S.; Colpo, G. D.; Hinds, S. N.; Morales, D.;

Ahnstedt, H.; Paz, A. S.; Assing, A.; Liu, F.; Juneja, S.; Kim, E.; Cho, S. M.; Gusdon, A. M.; Dash, P.; McCullough, L. D.; Choi, H. A., Brain injury, endothelial injury and inflammatory markers are elevated and express sex-specific alterations after COVID-19. *J Neuroinflammation* **2021**, *18* (1), 277.

Authors

Aden Geonhee Lee is currently a sophomore at Phillips Exeter Academy. He is passionate about neuroscience and enjoys scientific experimentation, especially computational and cellular biology.

DOI: 10.36838/v4i4.1

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

A New Method of Measuring Total Polar Compounds in Frying Oil: RGB Color Code

Boran Duman
American Collegiate Institute, Göztepe İnönü St. No:476, Konak, İzmir, 35290, Turkey; boranduman2005@gmail.com

ABSTRACT: During the frying process, physical and chemical changes occur within frying oil and produce harmful components named Total Polar Compounds (TPC) which also threaten human health. TPC measuring methods are generally expensive, time-consuming, and require special equipment. In this study, we aimed to develop a simpler, more accessible, and affordable method to measure TPC in frying oils by taking Red-Green-Blue color codes (RGB) as a basis. Waste oil was diluted with bottled oil into 10 different concentrations. Samples' TPC values were measured with a Testo 270 device and corroborated with a spectrophotometric measuring method. Additionally, the samples' pictures were taken and RGB's of individual samples have been noted with the help of a computer and cell phone. It is indicated that there is a direct correlation between acidity and TPC. RGBs have a high correlation with TPC. TPC level which is measured by RGBs on images gathered from both the computer and cell phone are found compatible with TPC reference values. Best results are obtained from the Red/Green ratio. Through this quick, easy, inexpensive, repeatable method; the TPC content of frying oil can be measured easily with the help of a photograph.

KEYWORDS: Chemistry; Organic Chemistry; Frying oil; Total Polar Compounds; Color Code.

Introduction

Frying is one of the most popular cooking processes in the world and it is defined as cooking of food items in (150-190ºC) oil.[1] In this process heat and mass transmission react together. During the transmission of heat from oil through food items, water escapes and it is replaced by oil. This process causes chemical and physical reactions to take place in the oil. These reactions induce a change in viscosity, darkening of color, foaming, and a change in the oil's smoke point.

Frying oils which are exposed to high temperatures for a long time and used repeatedly form various harmful compounds for human and environmental health.[2,3] The process causes three fundamental reactions: hydrolysis, oxidation, and thermal decay.[4] As a result of all these reactions, a large number of volatile and non-volatile polymerization products are formed.[5] Non-volatile polar substances contain fat-soluble and suspended substances, and they are named as Total Polar Compounds (TPC). There are more than 400 kinds of degradation products including free fatty acids, monoglycerides, diglycerides, sterols, TAG polymers, carotenoids, antioxidants, and soaps. The amount and chemical compositions of these substances depend on multiple factors such as type of nutrient and the oil, temperature, frying duration, and method of frying. However, the common feature of these degradation products is that their chemical structure is polar. Therefore, by measuring the number of polar substances in frying oil; in other words, by indicating the total polar component content (%TPC), it is feasible to make an evaluation of the total content of degradation reactions. These variables are crucial indicators for the quality and healthiness of the oil.[6]

According to multiple studies, the polar substances formed in frying oils have many negative effects on health.[2,7,8] Impacts of polar substances on human health vary depending on the type of the oil and polar component.[8] Polar substances can alter carbohydrate, protein, and lipid metabolisms. Thus, they create toxicological, cytotoxic effects, and their redox potential can modulate energy metabolism. This may cause various diseases (Cancer, Coronary Artery Disease).[8,9] According to International Food Law and Policy, frying oil gets into the status of waste oil and becomes inconvenient for cooking when the TPC content hits 25%.[10,11] In terms of TPC amount in frying oils, governments have felt it necessary to set restrictions on them because of the potential health effects. In The European Union, the maximum (TPC) amount has been set between 24-26% by experts. This limit value is 24% in Germany; 25% in Belgium, France, Portugal, Italy, Turkey and Spain; 27% in Australia, China, Switzerland.[2,12,13]

There are some techniques used to determine the amount of TPC in frying oils. In this context, the Official Analytical Chemists Association and the International Standards Organizations' methods are used as a reference.[14-16] However, these techniques are time-consuming, expensive, require expertise and lab conditions, and especially require routine monitoring of commercial and industrial frying oils methods that involve sample preparation procedures that make it difficult.[17] Thus, in recent years, methods and devices have been developed which are portable and detect polar substance content by using a dielectricity method or taking biosensory data as the basis. Some of the most popular rapid tests in order to evaluate frying oils utilize the changes in the permittivity of the oil which depends on the level of polar compounds in it. An ideal rapid test should be low cost, easy to use, practical, repeatable, give precise results and not require calibration. Additionally, supplemental chemical substances should not be needed, and the oils' temperature should not need to be

DOI: 10.36838/v4i4.2

adjusted before taking measurements.

Nowadays, there are three different methods (Fri-check, Testo 270, and Viscofrit) of measuring TPC rapidly and directly based on physical changes in the frying oil and TPC amount.[18-20] Firstly, Fri-check measurements take the oil density, viscosity, and interface based on changes in voltage. The method is reliable and shows a correlation with the results that are gathered from reference measurements.[18,21,22] Secondly, the Testo 270 device measurements take changes in the dielectric constant of the frying oil as a basis in its working procedure. Measurements that are taken by Testo 270 have shown good correlations in terms of frying duration and polar compound percentages. Among the fast methods that are examined, Testo 270 gives the most accurate measurement. Therefore, it can be used as the reference method. Meticulous calibration of the equipment is required. Meanwhile, it's suggested to use a calibration oil with a known source in order to set the device. Thirdly, Viscofrit measurements depend on changes of viscosity in the oil.[20] As long as procedures are followed correctly, this method is also reliable. However, special attention should be paid to calibration, cleaning, test temperature, oil filtering, and hygiene of the funnel.

All these tests require a certain temperature and a reference material to calculate accurately. Although they are repeatable, these tests are costly and have controversial accuracy. In this regard, there has still been a need for a faster and more cost-effective measurement system.

In chemistry, there are numerous measuring methods that are based on color, such as spectroscopy. As it is known that every matter has a color tone in nature, and these tones also have an equivalent value in a digital environment. In this digital coding system, RGB is the most known and commonly used color code system. This abbreviation stands for the three main colors. RGB; green, red, and blue. Every color tone has an equivalent numerical value. This coding system can be used for color tone-based measurements and can make subjective assessments of color tone objective. Thus, it will contribute to the accuracy and objectivity of the color-based measurement.

In this study, we aimed to develop a simple, affordable, and practical method for measuring TPC values of sunflower oil used for frying in order to make an RGB color-based analysis of the oil quality made by everybody everywhere, so that it will not require expertise.

Materials and Methods

Waste oil used in the experiments was gathered from a ministry authorized vegetable waste oil collection company, Habitat Recycling and Environment San. Tic. Ltd. Company.

Preparation of Frying Oils:

All experiments were carried out in Ege University Faculty of Medicine, Laboratory of the Department of Biochemistry. The oils with various TPC values were prepared as follows: Commercially purchased and unused oil is mixed with used waste oils in various proportions (0% - 100%) and standards are obtained. (Figure 1)

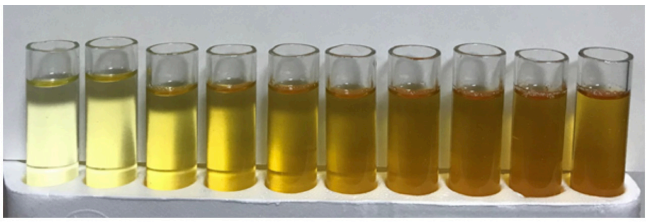


Figure 1: Outlook of samples for color analysis

Measurements Done by Testo 270:

A Testo 270 instrument was purchased for reference TPC measurements. TPC values in oil samples were measured by following the suggested application in the manufacturers' guide. Briefly, prepared oil standards were stored in 100 ml containers individually and were held in Benmari, 55℃ for 30 minutes. After that, TPC measurements were carried out with the help of the Testo 270 device, and values were recorded. Each concentration measurement was repeated 5 times, and results were indicated as the average of five measurements of each concentration. (Figure 2)

Figure 2: Samples In Benmari and Testo 279

Peroxide, acidity, and density indication:

Indication of peroxide, acidity, and density levels of samples were done in an accredited laboratory in Ege University Argefar Environment and Food Analysis Laboratory.

Spectrophotometric Measurements:

Samples taken from oils were prepared, dropped, and placed in a 96-hole bowl for spectroscopic inspection. Samples were scanned spectrophotometrically by Thermo Varioskan device (Thermo Fisher Scientific, Vantaa, Finland) between 300-760 nm, and the wavelength with the maximum absorbance ws indicated as 335 nm. Measurements were repeated 5 times for each sample.

Optic Measurements on Computer and Smartphone:

Examples prepared for spectrophotometric measurements were simultaneously placed in standard glass laboratory tubes and then photographs were taken with the help of a smartphone (Samsung Galaxy S9+) in order to analyze the color. (Figure 1 and 3)

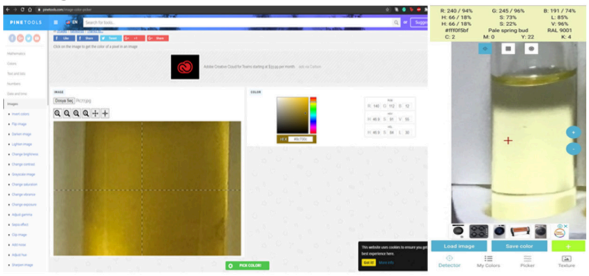


Figure 3: Color analysis applications for computer and smartphone

DOI: 10.36838/v4i4.2

RGB color codes were obtained by using PineTools (Online image color picker program on computer) and Color Detector (Color picker program on smartphone). RGB color codes were obtained from ten different points. Each picture measurement was repeated 5 times, and results were indicated as the average of five measurements of each concentration.

Measurements made with Digital Image processing and machine learning techniques:

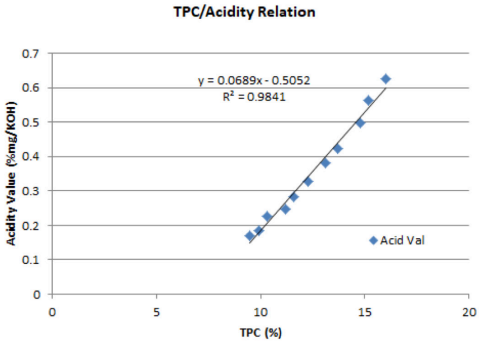
In order to conduct the color analysis of the photographs on a computer or another type of smart device, OpenCV library with the Python 3 program was used and machine learning techniques are applied. RGB values were obtained from standard 9 dots on every tube. These 9 measurement averages were taken into account for analysis. Additionally, the TPC ratios were calculated with the help of smart learning techniques regarding RGB / TPC values obtained on the computer.

Statistical Evaluation:

Statistical analyses were made in the Department of Medical Informatics and Biostatistics, Ege University, and IBM SPSS 25.0 program was used. Results that belonged to samples are given as averages and standard deviations. TPC measurements which were made with Testo 270 device utilized as the golden standard. TPC values that were calculated by spectroscopic measurement results, computer RGB color codes, smartphone RGB color codes, and the color codes calculated by the image processing technique were analyzed with the help of a compatibility test. Two-way mixed model and absolute agreement type were chosen for Reliability Analysis Statistics.

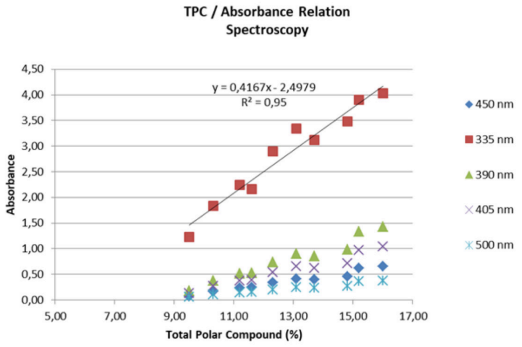
Results

TPC value of waste oil, which was utilized in the experiment, was 16%, the acid value was 0.626% mg / KOH, density was 0.914 g / ml and peroxide value was measured as 15 meq/kg. The mint state frying oil, which was bought from the market, TPC value was measured as 9.5%, acid value 0.169% mg / KOH, density 0.912 g / ml, and peroxide value was determined to be 12 meq/kg. When the waste oil was diluted with the mint state oil, it was clearly observable that acidity decreased. As the acidity value increases, the RGB color code changes. Oil's acidity was predicted by RGB color code as 96%. (Graph 1)

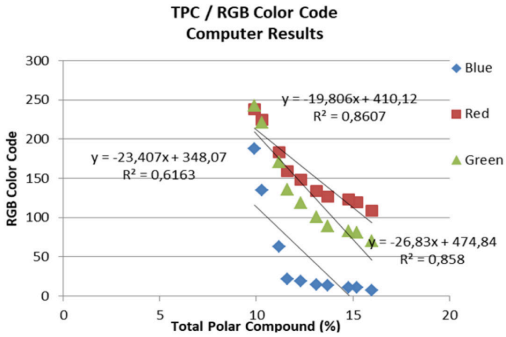


Graph 1: TPC/Acidity Relation

It was shown that TPC values that are measured with Testo 270 and values measured by spectrophotometer have a high correlation (Graph 2). Our results show that RGB color codes have a high correlation with TPC values. TPC values which are measured by Color codes of oils on images gathered from both the computer (Graph 3) and cell phones (Graph 4) are found compatible with TPC reference values which were measured by Testo 270 device.

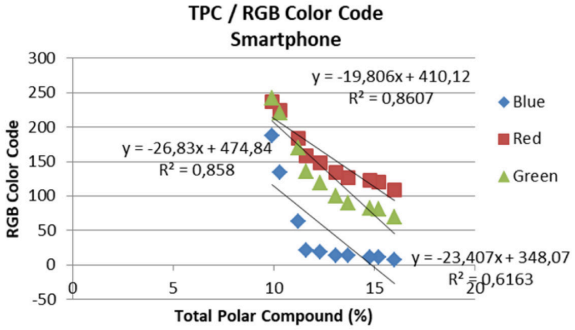


Graph 2: Scatter plot belong to TPC and different absorbance values



Graph 3: RGB color code measured with the help of a Computer and TPM relationship scatter plot.

Best results were obtained from the R/G ratio. Both computer and cell phone measurements were compatible with measurements gathered from machine learning techniques.



Graph 4: RGB color code was measured with the help of a smartphone and TPM relationship scatter plot.

According to the results of the reliability statistics, the nearest measurements to our gold standard, which is made with the Testo 270 device, were the results of the R/G ratio. Computer, smartphone, and machine learning calculations of R/G proportion have shown great correspondence to each other (Table 1).

Discussion

In this study, a new method of measuring TPC values of frying oils based on RGB color codes was developed. Since there

is no study done with RGB in the literature, our study brings innovation to oil measurement methods as a practical method based on RGB code. Additionally, it is shown in our study that the TPC value can be extrapolated from a picture that has been taken by a smartphone with the help of machine learning techniques. The operability of our method is tested with the help of basic software.

Table 1: TPC results obtained from different methods and correlation with Testo 270.

	Testo 270	Spectroscopy	Computer	Smartphone	Machine Learning
100%	16,00	15,59	15,63	15,71	15,61
90%	15,20	15,31	14,84	15,45	14,74
80%	14,80	14,34	14,95	14,44	15,10
70%	13,70	13,53	14,32	14,33	14,49
60%	13,10	14,03	13,38	13,30	13,40
50%	12,30	13,02	12,52	12,47	12,62
40%	11,60	11,33	11,72	11,16	11,55
30%	11,20	11,53	10,78	10,80	10,69
20%	10,30	10,60	10,14	10,23	10,05
10%	9,90	9,14	9,80	10,19	9,84
mean	12,81	12,841	12,,809	12,808	12,,809
variance	4,42	4,49	4,53	4,55	4,72
correlation with Testo270		0,967	0,987	0,985	0,981

Generally, in frying oil TPC measuring methods, it is required to heat the oil, to calibrate the tool with reference oil at regular intervals, to use some chemical substances for color comparison, or to reference a color catalog before analysis.[22] None of these are practical. As stated in our study, it is feasible to calculate the TPC value of a particular frying oil with the help of only one picture taken by a smartphone. The mechanism is easy, practical, repeatable, and affordable. In this regard, the main question of our study has been answered. Our method is tested with the TPC measuring device that has already been used and spectrophotometric measurement methods.

According to the results of the compliance statistics, the nearest measurements to our gold standard, which is made with the Testo 270 device, were the results of the R/G proportion. PC, smartphone, and machine learning calculations of R/G proportion have shown great relevance to each other.

There aren't adequate studies regarding the usage of image analysis to define the oil quality. RGB color codes can be used in defining the type of vegetable oils.[23,24] Their cooking quality can also be evaluated in another study by making a color analysis of images of the food items after they are cooked.[25] In another study, it may be possible to indicate the rottenness of the oil using image processing techniques, artificial neural networks, and evaluating RGB values.[26] Besides our main focus, one of the major indicators of the oil quality is acidity value, and it is also feasible to calculate that using RGB color codes.

In our study, we have demonstrated that TPC values in frying oils can be measured by image processing. More recently in another study, it is shown that nucleic acid assay can be measured via RGB color code analysis with the help of a smartphone application.[27] This contributes to the reliability of our study.

We determined that the TPM value of the waste frying oil we procured from the waste oil collection center is 16%. This shows that while trying to be extremely sensitive about health, unnecessary disposal of waste frying oil is happening. Hence, this dilemma has a negative impact on both the country's economy and the environment. Thanks to the method we have found, the amount of TPC can be determined with a photo anywhere. Thus, the preservation of the environment won't be forgotten while providing an economic profit.

Conclusion

It will be feasible to directly determine when the oil will be useless and harmful with this quick, easy, affordable, repeatable method. Before frying oils reach a state that is harmful to human health, restriction of their usage at the right time will be easily decided by applying this method. Through this technique, the TPC content of the oil can be measured easily with the help of a photograph. Furthermore the harmful effects of oil on human health will be prevented. Additionally, it will contribute to the renewal of the environment and state economy.

Acknowledgements

During this project, thanks a lot to Erhan Canbay and Zihni Onur Uygun from Ege University Faculty of Medicine, Department of Medical Biochemistry for providing their laboratory in order to put my experiments into practice; thanks to Emre Sarı from Izmir Institute of Technology, Faculty of Science, Department of Photonics and to my advisor Oktay Ünal.

References

1. Choe, E.; Min, D. B. Chemistry of Deep-Fat Frying Oils. *Journal of Food Science* 2007, 72 (5).
2. Hosseini, H.; Ghorbani, M.; Meshginfar, N.; Mahoonak, A. S. A Review on Frying: Procedure, Fat, Deterioration Progress and Health Hazards. *Journal of the American Oil Chemists' Society* 2016, 93 (4), 445–466.
3. Rakıcıoğlu, N.; Baysal, A. Yağda Kızartma yöntemi ile Pişirmede Oluşan Fiziksel ve Kimyasal Değişiklikler ve bunların insan sağlığı üzerine Etkisi. https://beslenmevediyetdergisi.org/index.php/bdd/article/view/777 (accessed Jan 30, 2022).
4. Sumnu, S.G., & Sahin, S. Advances in Deep-Fat Frying of Foods. CRS Press 2008.
5. Depren, E, Seven, Ü., Güçer, Ş. (2008). Isıl İşlem Sırasında Zeytinyağında Meydana Gelen Fiziksel Ve Kimyasal Değişimler Paper presented at the I.Ulusal Zeytin Öğrenci Kongresi, EdremitBalıkesir.
6. https://lipidlibrary.aocs.org/chemistry/physics/frying-oils/determination-of-polar-compounds-in-used-frying-oils-and-fats-by-adsorption-chromatography(accessed Jan 30, 2022).
7. Innawong, B.; Mallikarjunan, P.; Marcy, J. E. The Determination of Frying Oil Quality Using a Chemosensory System. LWT - *Food Science and Technology* 2004, 37 (1), 35–41.
8. Yuan, L.; Jiang, F.; Cao, X.; Liu, Y.; Xu, Y.-J. Metabolomics Reveals the Toxicological Effects of Polar Compounds from Frying Palm Oil. *Food & Function* 2020, 11 (2), 1611–1623.
9. Li, J.; Li, X.; Cai, W.; Liu, Y. Comparison of Different Polar Compounds-Induced Cytotoxicity in Human Hepatocellular Carcinoma HepG2 Cells. *Lipids in Health and Disease* 2016, 15 (1).
10. Dobargarnes, M. C.; Márquez-Ruiz, G. *Grasas y Aceites* 1998, 49 (3-4), 331–335.
11. Paul, S.; Mittal, G. S.; Chinnan, M. S. Regulating the Use of Degraded Oil/Fat in Deep-Fat/Oil Food Frying. *Critical Reviews in Food Science and Nutrition* 1997, 37 (7), 635– 662

DOI: 10.36838/v4i4.2

12.Ben Hammouda, I.; Triki, M.; Matthäus, B.; Bouaziz, M. A Comparative Study on Formation of Polar Components, Fatty Acids and Sterols during Frying of Refined Olive Pomace Oil Pure and Its Blend Coconut Oil. *Journal of Agricultural and Food Chemistry* 2018, 66 (13), 3514–3523.
13.Firestone, D. Regulation of Frying Fat and Oil. *Deep Frying* 2007, 373–385.
14.AOAC. (2002). Official Method 982.27. Polar components in frying fats, Chapter 41.41. 34.
15.AOCS, (2020). Determination of Polar Compounds in Used Frying Oils and Fats by Adsorption Chromatography. Retrieved from https://lipidlibrary. aocs.org/chemistry/physics/fryingoils/determination-of-polarcompounds-in-used-frying-oils-and-fats-by-adsorptionc
16.AUN, J. A. M. E. S. K.; CANTRILL, R. I. C. H. A. R. D. Process for Development of Standard Methods for the Analysis of Fats, Oils and Lipids. *Advances in Lipid Methodology* 2012, 273–299.
17.Correia, A. C.; Dubreucq, E.; Ferreira-Dias, S.; Lecomte, J. Rapid Quantification of Polar Compounds in Thermo-Oxidized Oils by Hptlc-Densitometry. *European Journal of Lipid Science and Technology* 2014, 117 (3), 311–319.
18.Frais. (20 October 2009). Fri-check. [WWW page]. URL. http//:www.frais.com.br/.
19.Ana Sayfa.https://www.testo.com/tr-TR/services/video-trainings-testo-270 (accessed Jan 30, 2022).
20.https://host.io/viscofrit.com (accessed Jan 30, 2022).
21.Gertz, C. Chemical and Physical Parameters as Quality Indicators of Used Frying Fats. *European Journal of Lipid Science and Technology* 2000, 102 (8-9), 566–572.
22.Osawa, C. C., Gonçalves, L. A. G., & Grimaldi, R. (2005). Nova ferramenta destinada ao monitoramento e à inspeção do descarte "in situ" de óleos e gorduras de fritura. Revista Brasileira de Vigilância Sanitária, 1(2), 102e107.
23.Fengxia, S.; Dishun, Z.; Zhanming, Z. Determination of Oil Color by Image Analysis. *Journal of the American Oil Chemists' Society* 2001, 78 (7), 749–752.
24.Milanez, K. D.; Pontes, M. J. Classification of Edible Vegetable Oil Using Digital Image and Pattern Recognition Techniques. *Microchemical Journal* 2014, 113, 10–16.
25.Sabbaghi; Hassan; Ziaiifar; Mohammad, A. [pdf] color quality variation of french fries during frying using image processing: Semantic scholar. https://www.semanticscholar.org/paper/Color-quality-variation-of-french-fries-during-Sabbaghi-Hassan/e755a91488fada431b010e11772b18aa5a95215d (accessed Jan 30, 2022).
26.(pdf) quality identification of used cooking oil based on https://www.researchgate.net/publication/303840408_Quality_Identification_of_Used_Cooking_Oil_Based_on_Feature_Fusion_of_Gas_Sensor_and_Color (accessed Jan 30, 2022).
27.Aydın, H. B.; Cheema, J. A.; Ammanath, G.; Toklucu, C.; Yucel, M.; Özenler, S.; Palaniappan, A.; Liedberg, B.; Yildiz, U. H. Pixelated Colorimetric Nucleic Acid Assay. Talanta 2020, 209, 120581.

Authors

Boran Duman is a 16-years-old high school student who is interested in multiple branches of science and technology. He is excited to participate in environmental scientific research. His ambition for developing and inventing practical inventions and research methods directed him to this project.

IJHSR
ijhighschoolresearch.org
■ RESEARCH ARTICLE

Using Genetic Algorithms to Solve Computer Science Olympiad Optimization Problems

Can Almelek, Ozgur Guzeldereli
Uskudar American Academy - Selamiali, Vakıf Sk. No:1 D:No:1, 34664 Üsküdar/İstanbul, Turkey; canalmelek44@gmail.com,

ABSTRACT: Genetic algorithms are random search algorithms based on biological evolution theory. In this study, the utilization of genetic algorithms in solving optimization problems from informatics Olympiads is investigated. It is hypothesized that genetic algorithms will be able to propose solutions to the optimization problem within a predetermined percent margin of the actual solution. A genetic algorithm is implemented in the C# programming language to test this hypothesis. Three tests are conducted, and their data is examined with various parameters taken into consideration. Consequently, the implemented genetic algorithm has been demonstrated to successfully solve the varying number of inputs (500, 1000, 2000) within an acceptable range. It is also concluded that several improvements and optimizations must be made to utilize this algorithm in a competitive landscape as time and memory constraints are exceeded. Furthermore, the tests concede that the number of inputs and the number of generations required to converge on an optimal solution is directly correlated. Therefore, this outcome should be taken into consideration when designing and tuning genetic algorithms with various parameters if the problem is aimed to be solved within the time and memory constraints of the competitive landscape.

KEYWORDS: Systems Software; Algorithms; Competitive Programming; Genetic Algorithms; Optimization.

■ Introduction

Computer science has become the epitome of global development into a more modern and enhanced society. With countless advancements in the fields of Artificial Intelligence (AI), Virtual Reality (VR), Augmented Reality (AR), robotics technologies, etc.; computer science has validated its crucial role in the new age. Therefore, there is a constant endeavor to improve the effects of computer science by coming up with new methods as well as educating more people. To be precise, the percentage of high schools offering computer science courses in the US has increased from 35% to 51% in the last 3 years.[1] This is a testament to the fact that more education institutions, students, teachers, and individuals emphasize computer science and its potential in the future.

As computer science's popularity prospers, there have been numerous means to assess one's knowledge and skills on a computer. There are coding websites, mainly Codeforces and Hackerrank designed to impose problem-based coding with occasional contests in a competitive manner. This system of problem-based coding in a competitive manner is essentially called competitive programming. The broader rendition would be an Olympiad in Informatics on regional and international scales. The world's most prestigious computer science competition is the International Olympiad in Informatics (IOI), which attracted 327 contestants from 97 countries in IOI 2019 Baku, Azerbaijan; all of whom passed several stages in national contests.[2] In these contests, there are common problem types such as Graph Theory, Game Theory, and recursion which are regularly displayed in problem tasks. One of the key objectives is problems that require finding the best solution from all feasible solutions such as optimization problems. There are numerous algorithms designed to solve optimization problems with the motive of determining the most optimal solution. Genetic algorithms (GAs) are one of the most fundamental algorithms that incorporate biology and genetic classification into computer science as a viable solution method.[3]

This study demonstrates the utilization of genetic algorithms in solving optimization problems from the Olympiad in informatics. It is hypothesized that genetic algorithms would be able to propose a solution within a certain percent margin of accuracy to Olympiad problems which are conventionally solved by other means, given the correct parameters. Within this paper, the structure of genetic algorithms will be explained, optimization problems will be defined, our implementation of a genetic algorithm will be shown, and with the given implementation an optimization problem will be attempted to be solved.

■ Methods

Genetic Algorithms:

Genetic Algorithms are optimization algorithms that are inspired by natural selection. Genetic algorithms were first designed by John Holland at the University of Michigan in the early 1970s.[4] They involve processes such as crossover and mutation to optimize for a constraint. Genetic algorithms can be used to generate feasible and high-quality solutions for search and optimization problems. Although the structure of genetic algorithms can vary slightly with their implementation, most, if not all, genetic algorithms have several commonalities: initialization, selection, offspring generation, mutation, and termination.[5]

Initialization:

All genetic algorithms start with an initialization step. Genetic algorithms consist of a population of individuals.

DOI: 10.36838/v4i4.3

Individuals carry a gene sequence or DNA. The DNA of an individual represents a solution to the problem that is aimed to be solved and the DNA should be structured accordingly. Therefore, each individual is required to be initialized to ensure they have at least some DNA. Although many techniques could be applied to initialize the individuals, two techniques are most prevalent. Of these two methods, assigning random genes to individuals is the most common. The other method is seeding the individual which signifies giving random genes that could be precise to the optimal value of the system or the problem. Initializing the individuals is quintessential to a genetic algorithm. Randomly assigning gene values makes genetic-diversity possible within the population which is key to spanning the entirety of the solution set of a particular problem.

Selection:

After individuals are initialized, the selection process takes place. Besides a DNA sequence, the other metric every individual has is the fitness value of their DNA. The fitness value of an individual determines how optimal it is, or how close the individual is to an optimal solution relative to other individuals. The fitness value is calculated via a fitness function which is a problem-specific function determining the optimality of a solution that is proposed by the genes of an individual. Therefore, it is important to form a feasible fitness function and it is the hardest process that is involved in constructing a genetic algorithm. If the fitness function fails to produce high-quality fitness values, the genetic algorithm will not be able to create high-quality solutions to the problem. After a valid fitness function is created, every individual's fitness is calculated. Then the individuals are sorted in the population according to their fitnesses and a percentage of lowest fitness individuals is discarded. Some low-fitness individuals are left in the population to ensure the genetic diversity within the population.

Offspring generation:

After the selection process, a new empty generation is created and filled via the crossover of the previous generation's individuals. Crossover is the process in which the genes of an individual are mixed with another individual's genes to create offspring. The process involves selecting random points in the genes of the individuals. Then two empty children are created. The genes of the children are set to be the gene of one of the parents (one for each child). In the points that are randomly selected, the parent that the child is getting the gene from switches to the other parent thus crossing over. This process can be repeated multiple times to ensure more diverse results. Crossover's main purpose is to generate new offspring from high fitness individuals to further increase the fitness of the population. By selecting from high fitness individuals and crossing over their genes, a gene sequence with a higher fitness level is created.

Mutation:

As with crossover, the mutation is a genetic operator that is used in creating higher fitness solutions in a genetic algorithm. The mutation is the process of randomly changing the genes of the individual. There could be several consequences of this action. Firstly, a higher fitness solution could be reached which is the optimal case. Another case would be a lower fitness solution that could be reached in which case the mutated child will not be able to pass its genes to the next generation. Either way, genetic diversity is created within the population to ensure that the system doesn't converge to a local optimum. By randomly changing the genes of the children of the new generation, enough new genes are inputted into the system to widen the search space of the algorithm. In elitist genetic algorithms where the fittest individuals of each generation are passed directly to the next generation, the mutation is also excluded from elite individuals to protect the progress of the genetic algorithm as mutations can lead to a decrease in fitness.

Termination:

Termination is the last step of a genetic algorithm. The termination of the genetic algorithm may depend on several factors and several termination conditions. For example, the number of generations could be constrained as well as the time spent on the problem. Other constraints include but are not limited to reaching a certain fitness level, the solution satisfies the bare minimum requirements of the problem, or a point is reached where the successive generations no longer produce new results. Usually, after termination, a solution is outputted as a solution to the problem or optimization. Data from each generation can be used to plot data graphs and the evolution of the population can be analyzed to draw further conclusions.

Pseudocode of a Basic Genetic Algorithm:

1. Initialization - All individuals of the population are initialized with n-length genes which represent a solution to the problem.
2. Fitness - The fitness of each individual is calculated using a fitness function.
3. New generation - Follow the proceeding steps to create a new generation.
4. Selection - A percent of the population is selected. The higher the fitness the higher the chance of selection.
5. Crossover - From the selected individuals, randomly select 2 parents. Crossover their genes to make 2 new offspring.
6. Accepting - Place new offspring in a new population
7. Elitism - If the genetic algorithm is elitist, select the highest fitness individuals of the population and also place them into the new population.
8. Replace - Use the newly generated population for another run of the algorithm.
9. Test - Test for a termination condition. If the condition is met, terminate the algorithm and return the necessary data as a solution to the problem or optimization.
10. Repeat - Go to step 2.

Optimization Problems:

As mentioned previously, optimization problems are the problems that require finding the best solution out of all feasible solutions. The whole entity of feasible solutions is often referred to as sub-optimal (local) solutions.[6] A specific goal is asked to be met at an optimal value out of all sub-optimal solutions. These problems must provide the variable con

straints as specifications necessary for constructing a solution. Accordingly, programmers are generally obligated to find the minimum and maximum values of a function that satisfies the task's objective.

Moreover, optimization problems can be classified in two manners according to their variables: continuous or discrete. In the cases of discrete variables in an optimization problem, an integer, graph, or permutation must be identified out of a countable set. In discrete optimization, some or all of the variables in a model are required to belong to a discrete set; this is in contrast to continuous optimization in which the variables are allowed to take on any value within a range of values. The discrete set is mainly initialized as subsets, combinations, graphs, or sequences. On the other hand, the cases of continuous variables require a continuous function to be found. Continuous optimization means finding the minimum or maximum value of a function of one or many real variables, subject to constraints. The constraints usually take the form of equations or inequalities.

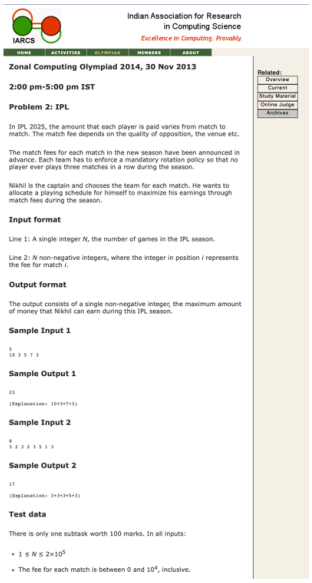
In general, a rather popular method for solving optimization problems is Dynamic Programming (DP).[7] DP is essentially a recursive solution with repeated method calls, storing the results of subproblems for the eventual solution. Compared to the aforementioned genetic algorithms, DP solutions usually use a lot of memory and have a longer running time. Even though genetic algorithms are typically harder to implement, efficiency in time and memory capacity are highly valuable assets for competitive programming which makes genetic algorithms more effective in Olympiad in Informatics.[8]

Results and Discussion

Genetic Algorithms can be considered systematic random search algorithms where the random search algorithm also considers the optimality of previous trials and evolves accordingly. Randomly initializing the individuals causes the proposed solutions to span the search space. With each generation, the solutions then attempt to converge at a global optimum. Mutations add genetic diversity to the gene pool and aim to ensure that the point the population converges is the global optimum. Knowing these it can still be inferred that a genetic algorithm may not be able to converge on the most optimal solution even with the best possible parameters as it is still a randomized and probabilistic algorithm. Although that may be the case, it could be expected that the final result of a genetic algorithm is close to the global optimum. For the set of runs that will be conducted in this paper, it is assumed that the problem is solved if the value of the solution proposed by the algorithm is within 5% of the actual solution after 2000 generations. The percentage values of how close the solution is to the actual solution are calculated using the following formula:

$$\left|\frac{proposed\ solution - actual\ solution}{actual\ solution}\right| \times 100$$

Another factor to consider is that genetic algorithms are intrinsically parallel compared to most algorithms which are serial. Serial algorithms only permit linear exploration that does not store every value reached by the algorithm's "infer" steps. On the contrary, genetic algorithms contain multiple offspring which allows multiple dimensional explorations. Depending on the accuracy of the offspring, some are eliminated, whereas some are pursued as promising solutions. Due to the parallelism, they can operate resembling a graph structure, evaluating multiple traceries at once. Thanks to their structure, particularly problems containing vast amounts of potential solutions are easier to solve which would take an immense amount of time with a simple search algorithm.


Indian Association for Research in Computing Science
Excellence in Computing. Provably.

Zonal Computing Olympiad 2014, 30 Nov 2013

2:00 pm-5:00 pm IST

Problem 2: IPL

In IPL 2025, the amount that each player is paid varies from match to match. The match fee depends on the quality of opposition, the venue etc.

The match fees for each match in the new season have been announced in advance. Each team has to enforce a mandatory rotation policy so that no player ever plays three matches in a row during the season.

Nikhil is the captain and chooses the team for each match. He wants to allocate a playing schedule for himself to maximize his earnings through match fees during the season.

Input format

Line 1: A single integer *N*, the number of games in the IPL season.

Line 2: *N* non-negative integers, where the integer in position *i* represents the fee for match *i*.

Output format

The output consists of a single non-negative integer, the maximum amount of money that Nikhil can earn during this IPL season.

Sample Input 1

5
10 3 5 7 3

Sample Output 1

23
(Explanation: 10+3+7+3)

Sample Input 2

8
3 2 2 2 3 5 1 3

Sample Output 2

17
(Explanation: 3+3+5+3)

Test data

There is only one subtask worth 100 marks. In all inputs:

- $1 \le N \le 2 \times 10^5$
- The fee for each match is between 0 and 10^4, inclusive.

Related:
Overview
Current
Study Material
Online Judge
Archives

Figure 1: ZCO 2014 Problem 4 IPL.[9]

Identifying the sections of a problem:

First and foremost, it is essential to identify crucial sections of a problem task before analyzing any algorithms or writing code. In this problem titled IPL, a certain number of games (N) are given according to the user's input. Afterward, N number of games is expected from the user with each containing a particular fee awarded if played in the game. The problem calls for the maximum amount of money that could be earned for an individual (Nikhil) through the N games. However, there is a condition in which an individual is obligated to play at most 2 consecutive games. Herewith, a certain algorithm must be constructed which maximizes salary while taking at most 2 consecutive game conditions into account (Figure 1).

Furthermore, a prediction about the complexity of a problem could be made per the given test data constraints. A maximum of $2*10^5$ number of games along with 10^4 for each game can be inputted by the user. An average computer is able to make approximately $2*10^6$ calculations per second regardless of the operating system. Thus, a brute force algorithm with time complexity of $O(n^2)$ would be sufficient for a solution to pass this problem. The time complexity of $O(n^2)$ signifies the worst solution possible for a problem which usually persists in iterating through every single value until the correct answer is reached. Nonetheless, a brute force solution wouldn't pass for this problem if the user could input $2*10^8$ number of games. In a hypothetical case, a genetic algorithm would be one of the most ideal solution methods for this problem with the time complexity of O(gnm). Among the time complexity of

DOI: 10.36838/v4i4.3

O(gnm), "g" stands for the number of generations, "n" stands for the population size, whereas "m" signifies the size of individuals.[10] In brief, although in this particular case it's not imperative, a genetic algorithm solution would be a fast and viable solution to the problem among random search and brute force algorithms.

Genetic Algorithms Implementation:

The genetic algorithm is implemented via the programming language C# which is a general-purpose, object-oriented programming language developed by Microsoft. The following section breaks down the code into several sections - namely initialization, selection, crossover, mutation, and termination.

Before the function of the code is analyzed, its inner structure must be understood.[11] First of all, the code starts with a class named Individual, which represents one individual in the entirety of the population. Each contains a set of genes that are of template type, a double fitness value, and a chromosome length which is common throughout the entire population.

```
class Individual<T> {
    public T[] genes; //instance variables
    private int chromosomeLength;
    public double fitness;
    ...
```

Its implementation is as follows:

All of these instance variables are initialized with a constructor as each instance of the Individual class is created. Another structure within the algorithm is the Population class. The class contains an array of individuals, a generation number, a population size, a chromosome size, and various probability values for the genetic algorithm to work on. The implementation of

```
class Population<T> {
    public Individual<T>[] individuals;
    public double generationNumber;
    public Individual<T> fittestMember;
    private int populationSize;
    private int chromosomeSize;
    private double mutationProbability;
    private Func<T> GetRandomGene;
    private Func<T[], double> CalculateFitness;
    private double elitismPercent;
    private double selectPercent;
    ...
```

its instance variables is as follows:

As evident from the code snippet, the population class also contains two functions: GetRandomGene() and CalculateFitness() functions. These functions are provided by the client of these classes and are called within the population's various functions. GetRandomGene() is the function that returns a random gene to initialize the members. CalculateFitness() is a function that returns a double value according to the genes of the individual; the number returned signifies the fitness of that particular gene set or individual. As cognizant of solely the meaning and implementation of these values, it is possible to analyze the code entirely. The first step in the code is the initialization of the individuals. The initialization method iterates through the array of individuals and sets the gene of each individual using the GetRandomGene() method. For the particular problem, the genes of the members are an array of integers. All values in the array are either 1 or 0. The gene sequence represents the sequence of games and each 1 in the gene represents that a match is going to be played while each 0 represents that the match will be passed.

After the initialization of the population is complete, the process of selection begins. Before the selection can order and remove individuals from the population, the fitness of each individual has to be calculated. The CalculateAllFitness() function located inside the Population class iterates through all of the population's individuals once more and calls the CalculateFitness() method on their gene sequence and sets their fitness value to be the return value of the method. The CalculateAllFitness() method is implemented as such:

```
public void CalculateAllFitness() {
    foreach (var individual in individuals) {
        individual.fitness = CalculateFitness(individual.genes);
    }
}
```

The CalculateFitness() function is provided by the client of the Population class. For this particular problem, the fitness is calculated to be the highest fee that can be reached within the gene sequence. A fee is calculated from a series of matches until 3 consecutive matches are played which is derived from the condition of the problem. If 3 consecutive matches are played, the fitness value is reset to 0 and the value it contains before resetting is compared to the *biggestFitness* value. If the fitness calculated from this particular series of matches within the gene sequence is bigger than the biggestFitness value, the *biggestFitness* value is set to be the fitness of this series. This process aims to maximize the fee earned from the gene sequence as it attempts to maximize for the largest value of fee possible while also keeping the amount of three consecutive matches played to the lowest possible value.

After all of the fitnesses are calculated the selection takes place. The selection function orders each Individual according to their fitness and a percent of the population whose fitness is lower is removed from the population. The percentage which the population is selected is given as a parameter to the method (selectPercent). The selection method is also a part of the Population class, and it is implemented in the following code snippet:

```
public void Selection(double selectPercent) {
    Individual<T>[] SortedList = individuals.OrderByDescending
                                    (o => o.fitness).ToArray();
    fittestMember = SortedList[0]; /* Fittest member will later be used to
                            collect data about the progress of the algorithm */
    individuals = SortedList.Take
                ((int)(populationSize * selectPercent)).ToArray();
}
```

After selection is completed, a new generation is created. Another array of individuals is initialized with all of the individuals being empty. Then a percentage of the previous generation is passed onto the next generation starting from the highest fitness. This is called elitism. The remaining empty places in the population are filled using the offspring of the individuals of the previous generation. To produce offspring two parents are selected using weighted random selection which uses the individuals' fitness as weight. After the selection is made, crossover and mutation are applied. The two offspring are placed into the new generation and this process is repeated until the new generation becomes just as populated as the pre

vious generation was. The code for making a new generation is implemented as follows:

```
public void MakeNewGeneration() {
    List<Individual<T>> currentGeneration = individuals.ToList();
    List<Individual<T>> newGeneration = new
                        List<Individual<T>>(populationSize);
    newGeneration = individuals.Take
            ((int)(elitismPercent * populationSize)).ToList(); /* Elitism is
  applied, a percent of the highest fitness individuals are passed onto the next
                                                              generation */

    Individual<T> parent1, parent2;
    double totalFitness = 0;
    foreach (var ind in currentGeneration){
        totalFitness += ind.fitness; /* Total fitness is calculated to weight
                                                    selection of new parents */
    }
    while(newGeneration.Count < populationSize) {
        parent1 = GetRandomByFitness(totalFitness); /* Weighted random selection
          is applied; higher the fitness, higher the chance of being selected */
        parent2 = GetRandomByFitness(totalFitness);
        var children = parent1.CrossoverReproduction(parent2);
        foreach (var child in children) {
            child.Mutate(mutationProbability, GetRandomGene);
            if (newGeneration.Count < populationSize)
                newGeneration.Add(child);
        }
    }
    individuals = newGeneration.ToArray();
}
```

All of these functions are put together in a PropagateGeneration method to simplify the entire process. PropagateGeneration is defined in the following way:

```
public void PropagateGeneration() {
    CalculateAllFitness();
    Selection(selectPercent);
    MakeNewGeneration();
}
```

The last step is the termination. For this particular problem, the termination condition is the number of generations that the algorithm ran for. In the tests made, every algorithm was run for 2000 generations before being terminated. The fitness levels were not adjusted or altered and were taken as is from when the algorithm stopped.

Evaluation of the Solution:

A total of three runs were made. Each algorithm is left to run for 2000 generations and with a population of 10000 individuals. Runs were made using N values of 500, 1000, and 2000. In these tests, fitness also represents the proposed solution of a gene sequence which in this case is the highest fee an individual can earn. This works as both the fitness and the solution to the problem are trying to be maximized. In each generation, 50 percent of the low fitness population is removed, and others are allowed to produce offspring. The mutation chance is 80%. This value can alter with the rate of convergence and prevent it dramatically, but it is required to be extremely high to cover the vast search space. Elitism percent is 5% meaning in each generation's highest fitness 5% of the population is passed onto the next generation without being subjected to any alteration. The processed data from the runs are entered into Table 1. The best fitness for each generation is plotted as a function of generation count and Figures 2, 3, and 4 are created.

The data suggests that for the tests done, the algorithm is successful in finding an optimal solution to the problem. It should be noted that for the third test the percentage difference is very close to the threshold value. This increase in the percentage difference is caused by inadequate parameters in the genetic algorithm, most notably the generation count and population size. Given a higher generation count, it is expected that the genetic algorithm converges closer to the value, resulting in a lower percentage difference. Another test with the same parameters except for generation count as test 3 is done to show this phenomenon. Note that the actual solution is different from test 3 as all the input values are randomized.

Solution Qualities	TEST I	TEST II	TEST III
Generation Count:	2000	2000	2000
Population Size:	10000	10000	10000
Inputs	N = 500	N = 1000	N = 2000
Best Fitness	1920717	3794093	7243546
Actual Solution	1950445	3874387	7614676
Percentage Difference	1.5%	2.1%	4.8%
Status	Solved	Solved	Solved

As can be observed from this test (Table 2), the percentage difference drops significantly when a larger generation count is used. Considering all of these tests, it could be concluded that genetic algorithms are adequate in solving such problems but require a lot of time and memory to run efficiently. Along with this, as the amount of input values gets bigger, the amount of generation required to converge on an optimal solution increases. This fact must then be taken into consideration when designing problem-specific genetic algorithms. Since time and memory are usually restrained on Olympiad questions, an efficient algorithm must be implemented with the best parameters possible.

Table 2: ZCO 2014 Problem 4 test II.

	Generation Count	Best Fitness	Actual solution	Percentage Difference
TEST III (Revised)	4000	7440479	7615239	2.3%

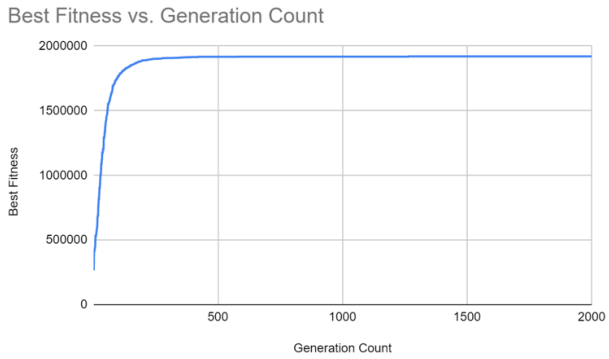


Figure 2: Best Fitness vs Generation Count Graph for Test I.

DOI: 10.36838/v4i4.3

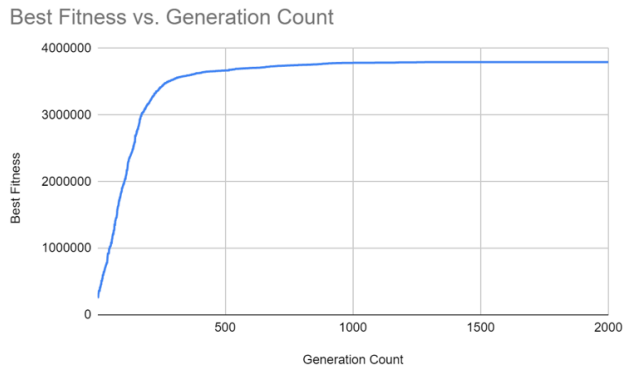


Figure 3: Best Fitness vs Generation Count Graph for Test II.

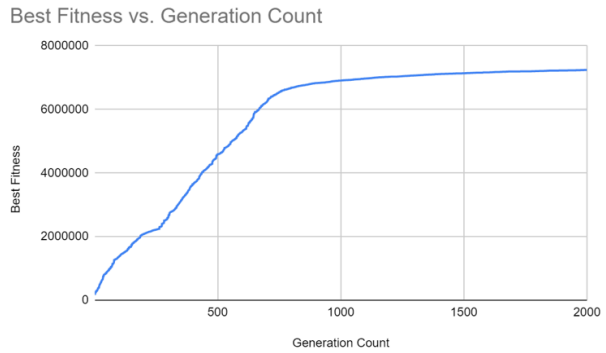


Figure 4: Best Fitness vs Generation Count Graph for Test III.

Conclusion

Thanks to the nature of genetic algorithms, competitive programmers and programming enthusiasts have the upper hand in solving optimization problems. Although genetic algorithms are able to provide optimal solutions for a majority of optimization problems, a couple of drawbacks are noted from the solution of IPL. Firstly, since genetic algorithms require a set of generations to be processed in order to reach the final accurate answer, time and memory constraints are generally exceeded. For instance, in the IPL example, the time limit is stated as 1 second, whereas the memory limit is stated as 32 Megabytes (MB). The solution that is at hand utilizing genetic algorithms requires at least 370 seconds and encompasses 200 MB of memory. Thus, certain alterations to the program should be made if genetic algorithms are to be used in a competitive environment. A suggestion for an alteration would be to stop new generation formation when an accurate solution is reached. A genetic algorithm provides a comprehensive search methodology for optimization. The problem of finding an accurate solution in a space with many sub-optimal solutions is a classic problem for all systems that can adapt and learn.

In conclusion, there have been benefits and drawbacks that have emerged as genetic algorithms were utilized in this study in order to solve optimization problems from informatics Olympiads. It is observed that genetic algorithms are generally able to solve such problems but are unable to comply with the memory and time restrictions. It should also be noted that converging on higher input values, requires more generations to be run on the algorithm, or increased population size. Consequently, a competitor in such informatics Olympiads must consider the possibility of using genetic algorithms as a tool for solving the problems yet must also approach the algorithm with caution as the implementation has to be optimized heavily in order to comply with time and memory constraints of the Olympiads.

Acknowledgements

Our IB High-Level (HL) Computer Science (CS)teacher Özgür Çiçek encouraged us to research genetic algorithms since it has been the IB CS HL Paper 3 topic for the past 2 years.

References

1. Klein, A. More than half of high schools now offer comput er science, but inequities persist.https://www.edweek.org/ teaching-learning/more-than-half-of-high-schools-now-offer-computer-science-but-inequities-persist/2021/11 (accessed Jan 7, 2022).
2. IOI Statistics. https://stats.ioinformatics.org/olympiads/2019 (ac cessed Jan 10, 2022).
3. Techopedia. What is a genetic algorithm? - definition from Te chopedia https://www.techopedia.com/definition/17137/genet ic-algorithm (accessed Jan 10, 2022).
4. Sablier. Coding games and programming challenges to code better https://www.codingame.com/playgrounds/334/genetic-algo rithms/history (accessed Jan 30, 2022).
5. Worden, K.; Liu, S.; Das, D.; Avi Ceder, A. Genetic Algorithm https://www.sciencedirect.com/topics/engineering/genetic-algo rithm (accessed Jan 11, 2022).
6. Lecture12-dukecomputerscience. https://www.cs.duke.edu/cours es/fall05/cps130/lectures/skiena.lectures/lecture8.pdf (accessed Jan 12, 2022).
7. Dynamic Programming https://www.geeksforgeeks.org/dynam ic-programming/ (accessed Jan 30, 2022).
8. Continuous optimization. https://uwater loo.ca/combinatorics-and-optimization/research-combinator ics-and-optimization/research-areas/continuous-optimization (ac cessed Jan 12, 2022).
9. Indian Computing Olympiad Archives 2012 - IARCS. https:// www.iarcs.org.in/inoi/2014/zco2014/zco2014-2b.php (accessed Jan 17, 2022).
10. Time complexity of genetic algorithms exponentially ... https:// fernandolobo.info/p/gecco00.pdf (accessed Jan 13, 2022). Agnishom. ZCOSolutions/ipl.cpp at master · Agnishom/Zcosolu tions.
11. Bajpai, P.; Kumar, M. Genetic Algorithm - an Approach to Solve Global Optimization Problems https://www.researchgate.net/ (ac cessed Jan 20, 2022).

Authors

Özgür Güzeldereli is a 17-year-old sophomore student at Usku dar American Academy in Turkey. He is interested in computer programming, theoretical computer science, and algorithms. He specializes in operating system development.

Can Almelek is a 16-year-old sophomore student at Uskudar American Academy in Istanbul, Turkey. His fields of interests in computer science entail competitive programming, algorithm analysis, artificial intelligence and object-oriented programming. He's also fond of visual arts, table tennis, basketball, and public speaking.

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

An Analysis on the Possibility of Establishing a Threshold Energy Level for the Solar Flares on the Interactions of Hard X-Ray Beams and the Ionosphere

Can Erol, Evren Sanlı, Defne Lara Balsarı, Muhsin Erhan
American Collegiate Institute, Göztepe İnönü St. No:476, Konak, İzmir, 35290, Turkey; canerol2411@gmail.com

ABSTRACT: This paper presents an analysis of the interactions of the hard x-rays emitted through the high-energy solar flares, often seen in the X and Y class, within the ionosphere which is located in the thermosphere. Hence, the paper concentrates on whether it's possible to specify a critical threshold energy level for the hard x-ray emissions from a solar flare that cause the x-ray beam to reach the Earth's crust and its inhabitants, assuming that the intensity of the hard x-ray beam is kept as a constant. This should serve as a critical point in determining the harmful effects of the solar flares in the atmosphere and furthermore on Earth as a whole. The methodology that has been followed is the application of a well-known scientific effect to the problem, the photoelectric effect, and the evaluation of the implications of its conclusions. It is then seen that as the intensity of the X-ray beam is kept as a constant, the energy of the incoming x-ray beam does not affect the process of atmospheric absorption, since energy is quantized by the photons and the decreasing wavelength does not contribute to the number of electrons liberated in the ionospheric D region.

KEYWORDS: Physics and Astronomy; Quantum Physics; Solar Flares; Ionosphere; X-Ray Emission.

Introduction

This research paper is focused on the examination of the mutual correlation between the hard x-rays emitted through the X and Y class high-energy solar flares with the ionosphere which is an integral part of the thermosphere.[1] The question we as a team specified our research into is "Can there be a critical energy threshold for the hard X-rays emitted through the solar flares in which the Earth's atmosphere cannot process it and if so, what would it be?" Our hypothesis was that the threshold could be found through the calculation and understanding of the Photoelectric effect that is naturally occurring by the interactions between the hard X-rays and the competence of the Earth's atmosphere and surface.[2]

For reference data, we researched the November 4, 2003, Giant Solar flare, it was the largest one currently recorded by scientists.[3] In essence, solar flares are incidents that have immense effects on both the natural species of our planet and the infrastructure of telecommunication technology. After thoroughly examining the example of the November 4, 2003, Giant Solar flare, we started to research a joint threshold that can define the amount of radiation the Earth's atmosphere could absorb. Solar flares emit many kinds of waves ranging from Gamma to X-Rays; however, for the ease of comprehension and drawing the limits of our paper, we narrowed our research down to the examination of the amount of X-Rays that solar flares emit only.[3]

To better understand how X-rays interact with the structure of the Earth's atmosphere, we firstly gathered information on the Earth's atmospheric layers, namely the thermosphere and its integral part the ionosphere, as shown in Figure 1. These two layers' main function is to absorb any rays that enter the atmosphere with the abundance of Oxygen and Nitrogen

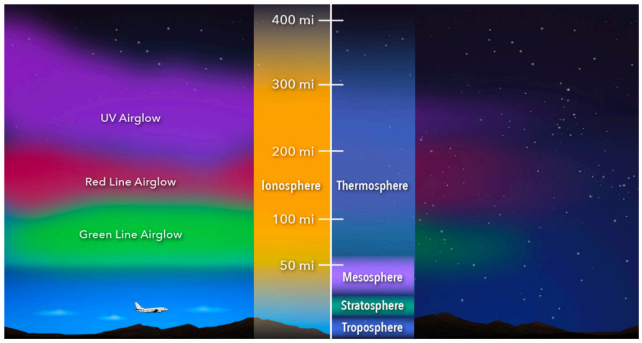


Figure 1: Diagram showing the layers of the Earth's atmosphere and ionosphere.[4]

gasses in their compositions.[5] Essentially, the thermosphere is the layer of the atmosphere which lies just above the mesosphere. Within this layer of the atmosphere, UV light induces photoionization of molecules, resulting in the formation of ions.[5] The formation of the ions further translates into the subsequent formation of the ionosphere. The ionosphere is a dense layer of electrons and ionized atoms and molecules that extends from 48 kilometers above the surface to 965 kilometers, intersecting with the previously mentioned layers of the mesosphere and thermosphere.[6] This mobile zone expands and contracts in response to solar circumstances, and it is subdivided into three regions: D, E, and F, depending on the wavelength of solar light absorbed.[6] Throughout the paper, the reader will come across the D region the most, mainly because this is the region where hard X-Ray beams interact with the atmospheric composition and are absorbed.[7] Since the ionosphere is composed of charged particles, it is extremely sensitive to changes in magnetic and electric conditions in space. These circumstances, along with other phenomena like bursts of charged particles, are referred

DOI: 10.36838/v4i4.4

to as space weather and are frequently associated with solar activity. The ionosphere is an essential connection in the Sun-Earth interaction chain, since through the ionosphere radio communications can be possible.[7] This paper portrays an analysis of a sophisticated and detailed research question regarding the X- Ray absorption abilities of our atmosphere that hasn't been thoroughly discussed by previous literature before.

Results and Discussion

The paper consists of two sections: Limitations of the Photoelectric Absorption and The Effect of Quantized Energy on the Amount of Emitted Electrons.

Limitations of the Photoelectric Absorption:

The photoelectric effect is the starting point of quantum mechanics in its modern form, and the idea of quantized energy. Hence, understanding the limitations of the photoelectric effect and the absorption of a photon by a material is crucial in evaluating the interactions between the x-rays and the molecules in the D region of the ionosphere. X-rays, and their carriers, photons, are by no means affected by the magnetosphere of the Earth, since they are not electrically charged. Thus, the only significant interaction between the incoming X-rays and the atmosphere occurs in the D region of the ionosphere which is located at the thermosphere of the atmosphere.[7] The incoming X-rays often collide with the atoms, and their electrons, located in the D region which is highly famous with the high density of electrons it includes. Hence, as the X-rays collide with N_2 and O_2 atoms, the energy coming from the X-rays are absorbed by the atom, and as a result, an electron is set free with a specific amount of kinetic energy. This is known as the photoelectric effect, and hence is important in evaluating the hypothesis since the increasing energy of the incoming photons does not specifically cause an increase in the emission of the electrons of a molecule after a certain frequency threshold.[8] However, it should be noted that the intensity of the incoming X-rays are to be taken into account as a constant, since the increase in the intensity of the X-rays is linearly correlated with the amount of electrons emitted.[8]

The Effect of Quantized Energy on the Amount of Emitted Electrons:

The calculation of the energy of an emitted electron via the photoelectric effect is made by the equation:[9]

$$K = hf - \Phi$$

In which K is the maximum kinetic energy that the electron can possess after being set free from the atom of the material, h is the Planck constant, f is the frequency of the incoming radiation, and Φ is the work function, which denotes the required energy needed to tear off an electron from the specified atom. Work function is a function of the material, and is higher in nonmetals compared to metals.[9] Recalling the approved postulations of the photoelectric effect, one can see that the quantized energy carried by singular photons can be only used in the tearing off of one electron and that after a threshold frequency, which is also a function of the specified material, the value of the frequency, and thus, the wavelength of the specified radiation is not accounted regarding the number of electrons emitted from a set of materials. The excess energy after subtracting the work function from the incoming photon, energy only affects the kinetic energy of the emitted electrons, and hence does not have any significant effect on the number of liberated electrons, as shown in Figure 2.

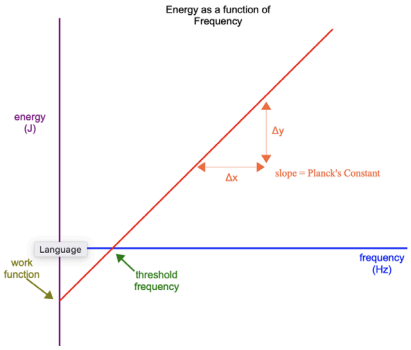


Figure 2: Energy as a function of frequency.[2]

This is indeed crucial in our understanding of the hypothesis and the x-ray interactions with the ionosphere. In the hypothesis we established, we consider the energy of an incoming beam, and we are trying to determine whether it is possible for the incoming x-rays to reach the Earth's crust and its inhabitants after a critical threshold energy level. However, our hypothesis does not specifically account for the intensity of the incoming beam, since it is also a major observational problem that is yet to be resolved by the scientific community. The minor interactions of the X-ray beams with the outermost parts of the atmosphere cause the diffraction and scattering of beams before entering the ionosphere and therefore prevent us from setting a specific conclusion regarding a critical threshold energy value for a solar flare that could be then seen as a danger threshold for the life on Earth. Briefly, the intensity of the incoming X-ray beams cannot be specified and vary significantly from their predicted source values. At this point, it is needless to say that the hard x-rays that this research concentrates on have wavelengths high above the threshold frequency needed to enable the photoelectric effect on the N and O molecules of the atmosphere.[10]

Conclusion

The research we have conducted over the course of the last two months allowed us to examine the interaction between the X-rays emitted through the solar flare activities and respected layers of the atmosphere of the Earth, namely the thermosphere and the ionosphere, especially the ionospheric D region. In the end, we have reached the conclusion that it is highly unlikely to determine a critical threshold energy level for the solar flares that enable the hard X-ray beams emitted from them to reach the Earth's crust, due to the complex diffraction and scattering patterns of the X-ray beams and the fact that as the intensity of the beam is kept as constant, the amount of energy a beam possesses does not affect the number of electrons ionized. Hence, it should be noted that further specifications and enhancements could be made by rigorous mathematical analyses.

Acknowledgements

We would like to thank our physics teacher Muhsin Erhan for his endless support to find our inspiration and motivation for the paper we successfully completed.

References

1. Tandberg-Hanssen, E.; Emslie, A. G. *The Physics of Solar Flares*; Cambridge University Press: Cambridge, 2009.
2. Kontar, E. P.; MacKinnon, A. L.; Schwartz, R. A.; Brown, J. C. Compton backscattered and primary X-rays from solar flares: An gle Dependent Green's function correction for photospheric albe do. https://www.aanda.org/articles/aa/abs/2006/06/aa3672-05/aa3672-05.html (accessed Apr 17, 2022).
3. Kane, S. R.; McTiernan, J. M.; Hurley, K. Multispacecraft observations of the hard X-ray emission from the giant solar flare on 2003 November 4. https://ui.adsabs.harvard.edu/abs/2005A%26A...433.1133K/abstract (accessed Apr 17, 2022).
4. GMS: Ionosphere Graphics. https://svs.gsfc.nasa.gov/12960 (accessed Apr 17, 2022).
5. Center for Science Education. https://scied.ucar.edu/learning-zone/atmosphere/thermosphere (accessed Apr 17, 2022).
6. HOLLINGWORTH, J. Structure of the Ionosphere. Nature 1934, 134 (3386), 462–462.
7. Sutherland, L. C.; Bass, H. E. Atmospheric Absorption in the Atmosphere up to 160 Km. *The Journal of the Acoustical Society of America* 2004, 115 (3), 1012–1032.
8. The Editors of Encyclopaedia. Photoelectric effect. https://www.britannica.com/science/photoelectric-effect (accessed Apr 17, 2022).
9. PRATT, R. H.; RON, A. K. I. V. A.; TSENG, H. K. Atomic Photoelectric Effect above 10 Kev. *Reviews of Modern Physic*s 1973, 45 (2), 273–325.
10. Watanabe, K. Ultraviolet Absorption Processes in the Upper Atmosphere. *Advances in Geophysics* 1958, 153–221.
11. Lesson 34: Photoelectric effect graphs - studyphysics. http://www.studyphysics.ca/2007/30/07_emr/34_photo_graphs.pdf (accessed Apr 16, 2022).
12. A parameterized model of x-ray ... - wiley online library. https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2018RS006666 (accessed Apr 16, 2022).
13. Švestka, Z. Solar Flares. 1976.

Authors

Can Erol is a student at American Collegiate Institute, Izmir, that is truly interested in physics and mathematics, and an avid learner of science looking forward to expanding his research interests. He currently wants to double major in physics and mathematics.

Evren Sanlı is a student at the American Collegiate Institute in Izmir, Turkey. He is deeply interested in physics, especially particle and astrophysics, and is hoping to widen his understanding in these fields through future research. Currently, he is planning to double major in mechanical engineering and econometrics.

DOI: 10.36838/v4i4.4

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

A Comparison of Production, Efficacy, and Safety of mRNA and Conventional Vaccines

Deniz Cenikli
Uskudar American Academy, Akasya Mah, Cecen Sok, Akasya Koru Sitesi, B3-C Girisi, Apt. No:46-G, Daire:183, Üsküdar/İstanbul, Turkey; dcenikli2005@gmail.com

ABSTRACT: This paper investigates the advantages and disadvantages of mRNA vaccines and conventional vaccines through a comparison of both, particularly in three areas: production, efficacy, and safety. Classical vaccines comprising of live-attenuated vaccines and inactivated vaccines. Live-attenuated vaccines use an attenuated version of the original pathogen whereas inactivated vaccines use inactivated germs. Messenger RNA vaccines, on the other hand, utilize mRNA containing the code for antigens. With this, the aim is to make cells build the antigens themselves. When these two are compared, it is seen that mRNA vaccines are faster to produce due to technological improvements. They are more effective than classical vaccines owing to the advancements in delivery systems. Moreover, they do not possess the risk of degradation of DNA or instability again due to technological development. On the other hand, inactivated vaccines are harder to produce because of developing cultures. They are not as effective as mRNA vaccines in SARS-CoV-2 and its mutations. Finally, live-attenuated vaccines have the probability of a reversion to the virulence of the natural germ. Thus, they cannot be given to people with weakened immunity. When all of these are considered, it has been concluded that mRNA vaccines have numerous advantages when compared to traditional vaccines.
KEYWORDS: conventional; inactivated; live-attenuated; mRNA; vaccines.

Introduction

With the COVID-19 pandemic, all the attention of the scientists was drawn to a new type of vaccine: mRNA vaccines. This was a new approach to vaccination since conventional vaccines consisting of live-attenuated and inactivated (dead) vaccines have used entire pathogens that are either weakened (attenuated) or inactivated (killed) through physical and chemical processes for producing vaccines. However, the research on mRNA vaccines, in fact, dates back to the 1960s. During that period, mRNA vaccines had some deficiencies that needed to be improved. Owing to technological advancement in recent years, this was made possible for mRNA vaccines, especially for mRNA COVID-19 vaccines.

This research paper will investigate the general properties of conventional vaccines and mRNA vaccines as well as compare the two vaccine types with respect to their production methods, efficiency for preventing diseases, and biosafety that they possess by focusing on SARS-CoV-2. The paper starts with an introduction of the two types of vaccines mentioning their properties and some of their benefits and disadvantages. The conventional vaccine classification was made according to the information taken from the National Institute of Allergy and Infectious Diseases. The introduction continues with the comparison of the vaccine types in three main areas which are manufacturing, effectiveness, and safety including elaboration on some additional convenience and inconvenience of the vaccines. The paper is finished by showing mRNA vaccines have several advantages over conventional vaccines with respect to biosafety, efficacy, and production.

Methods

Conventional Vaccines:

Conventional vaccines usually contain whole pathogens that are attenuated or inactivated so as to stimulate the immune system of the body against a pathogenic agent.[1] The first type of traditional vaccine is the live or attenuated vaccine. Attenuated vaccines are comprised of the weakened version of a virus or bacterium that causes the disease so that the germs replicate themselves several times and create an immune response in the body.[1] Most attenuated vaccines can create a strong and long-lasting immune response in the body against the disease with only one or two doses of the vaccine as the real pathogens and the pathogens utilized in this type of vaccine are very similar.[2] The virulence of these disease-causing viruses or bacteria is usually reduced by repeated culturing.[3] Since the pathogenic microorganism that is found in an attenuated vaccine is still a living organism even though it is weakened, live attenuated vaccines require refrigeration meaning that they need to be kept cool under low temperatures mostly between 2°C and 8°C (36 °F to 46 °F).[4] However, there are a few exceptions such as live varicella (chickenpox) vaccine which must be stored frozen within a temperature range of -50°C to -15°C (-58°F to 5°F) in a refrigerator.[4] Together the varicella (chickenpox) vaccine, measles, mumps, rubella (MMR) vaccine, shingles vaccine, and live attenuated influenza vaccine (LAIV) are some of the other examples of live vaccines.[5]

The second type of traditional vaccine is the inactivated or dead vaccine. In order to produce inactivated vaccines either

DOI: 10.36838/v4i4.5

entire pathogenic agents or microorganisms that are very similar to the germs are destroyed -or sometimes altered- through the agency of chemical agents, heat, or radiation after being grown in culture.[3] The aim of the destruction or alteration of the disease-causing microorganisms is to prohibit the replication of the replication of the disease-causing microorganisms.[6] A negative consequence of the destruction of germs is the decrease in the immunogenicity of inactivated viruses and bacteria provided by inactivated vaccines.[3] It is said that the immunity obtained from inactivated vaccines is generally shorter and weaker than the immunity provided by live attenuated vaccines.[3] Since the length and strength of the immune response of an inactivated vaccine are less, more doses of this kind of vaccine - generally two or three shots- are required to obtain a sufficient amount of antibodies in the body when compared to live-attenuated vaccines.[3] Although this may be a disadvantage of inactivated vaccines over live attenuated ones, an advantage of inactivated vaccines is the fact that the refrigeration process is not essential for some inactivated vaccines such as lyophilized vaccines in which the water is removed and they are vacuumed since the pathogenic microorganisms are inactivated through chemical or physical processes.[7] This eases the distribution of this type of vaccine around the globe as opposed to attenuated vaccines whose distribution is limited to the countries having refrigerators.[9] However, as usual, there are exceptions. Some of them are inactivated vaccines that are in a liquid state that do not contain aluminum adjuvants.[8] Additionally, inactivated influenza vaccine (IIV), hepatitis A vaccine, rabies vaccine, and polio vaccine can be given as other examples of inactivated vaccines that were developed until now.[9]

Messenger RNA Vaccines:

Messenger ribonucleic acid (mRNA) vaccines have been studied for decades dating back to the 1960s.[10] Recent advancements in technology in addition to the SARS-CoV-2 pandemic have extremely increased the speed of the development of mRNA vaccine technology. Messenger RNA is therefore a novel technology utilized to generate an immune response in the body.[1] Unlike classical vaccines, instead of entire pathogens, mRNA vaccines use mRNA in which the "blueprint" of a specific pathogen is encoded.[1] This allows the cells themselves to produce antigens (disease-causing proteins) which are normally found in the pathogenic bacteria or viruses by translating the information encoded in the mRNA.[1] Similar to live-attenuated vaccines, mRNA vaccines should be refrigerated. An exemplification of refrigeration can be given as the COVID-19 vaccine developed by Pfizer-BioNTech.[4] Since the durability and stability tests of the vaccine started quite late in the developmental process, refrigeration was imposed by the biotechnology company.[4] Likewise, another benefit of messenger RNA vaccines which is common to live-attenuated vaccines is their ability to generate strong immunogenicity, meaning that the body creates a strong immune response against the antigens that were produced by the body itself.[11] Yet, sometimes this immunogenicity might be exaggerated by the body causing the excess amount of cytokine which is a crucial and small protein used in cell signaling and produced by the innate immune system to appear in the body. However, there is no evidence supporting that "mRNA COVID-19 vaccines or even non-mRNA vaccines would result in cytokine storms".[12]

Two types of mRNA are currently being studied: non-replicating (conventional) and self-replicating (self-amplifying) mRNA.13 Non-replicating mRNA consists of a 5' cap structure that contains a 7-methylguanosine cap (m^7G cap) that is connected to the first nucleotide by a triphosphate bridge, 5' untranslated region (UTR), immunogen information, 3' untranslated region (UTR), and a 3' poly-A tail respectively from left to right as shown in Figure 1 (a).[13] On the other hand, self-replicating mRNA, different from non-replicating mRNA contains an RNA-dependent RNA polymerase (RDRP) complex between 5' UTR and immunogen code which is an enzyme catalyzing the replication of RNA -in this case, mRNA- in addition to the components found in the structure of non-replicating mRNA as also shown in Figure 1 (b).[13] In other words, the RDRP complex that is located in the self-amplifying mRNA enables the mRNA to multiplicate intracellularly as shown in Figure 2.[13] The 5' m^7G cap and 3' poly-A tail are used in both conventional and self-amplifying mRNAs to prevent the breakage of the immunogen blueprint by the enzymes found in the cytoplasm of the cells in the body.[13]

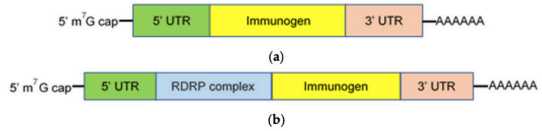


Figure 1: Non-replicating and self-replicating mRNA structures.[13]

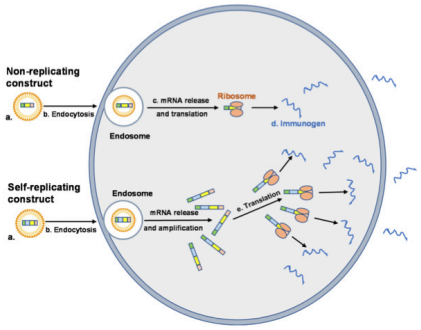


Figure 2: Production of immunogens via non-replicating and self-replicating mRNA.[13]

Production of Traditional and mRNA Vaccines:

Traditional vaccines take a longer time to be developed compared to messenger RNA vaccines.[1] In a nutshell, in order to produce live-attenuated vaccines first, wild bacteria or viruses are weakened under laboratory conditions.[3] The attenuation process might be done through heat or light although repeated culturing is the most preferred method.[3] Then, these viruses or bacteria are delivered to the body via live-attenuated vaccines. For inactivated vaccines, however, the methods such as using chemicals, radiation, or heat are used to destroy or kill the germs as can be seen in Figure 3.[14] After the inactivation, the destroyed pathogens are delivered to the body through inactivated vaccines Figure 3.[14]

DOI: 10.36838/v4i3.5

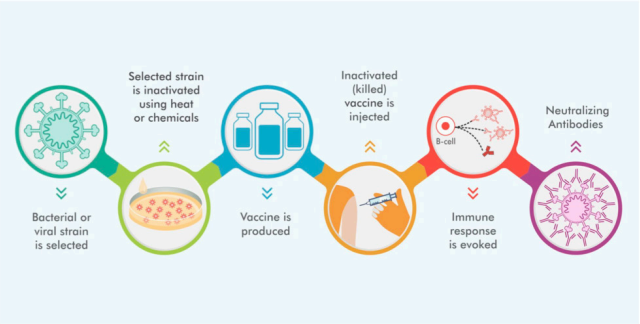


Figure 3: Production process of inactivated vaccines.[14]

Even though the description of the manufacturing processes of attenuated and inactivated vaccines seem to be short, indeed they are longer than the manufacture time of mRNA vaccines.[1] The average duration of vaccine development is measured to be 10 to 15 years.[15] One reason for this might be the fact that vaccine development comprises several stages: discovery and development, pre-clinical trials, human trials, licensing, manufacturing, and distribution.[16] To support this claim, the development times of several conventional vaccines from the discovery and development stage to the licensing stage (FDA approval stage) can be given. The vaccine for SARS-CoV-2 was developed in less than one year by breaking the fastest produced vaccine record.[17] In the meantime, the developmental time for the chickenpox (varicella) vaccine was 28 years, for the flu vaccine was 27 years, for polio was 13 years, and for mumps was 4 years.[16] Although the same vaccine development stages are applicable for mRNA vaccines as well, the time for the discovery and development stage is much shorter for them owing to the technological advancements in gene sequencing.[16] Specifically for COVID-19, the human trial stage was shortened as well as the discovery and development stage.[16] Another reason for the fact that traditional vaccines require a longer production time could be the culturing process in inactivated vaccines.[18] Since producing numerous inactivated bacteria and viruses takes up plenty of time, the discovery and development stage is longer for inactivated vaccines specifically, elongating the overall manufacture time for this type of classical vaccine.[18] Briefly, due to the differences in the manufacturing process, for the production of conventional vaccines, more time must be devoted.

The Manufacture and Working Processes of mRNA Vaccines:

Compared to traditional vaccines, mRNA vaccines have a different way of production and working as depicted in Figure 4.[19] First of all, the sequence of the pathogenic agent is identified via gene-sequencing.[1] Secondly, vaccines containing small and harmless mRNA fragments in which the code for the specific antigen -for COVID-19 this specific antigen is called spike protein- is encoded are produced and given to people.[1] Thirdly, mRNA fragments enter the body.[1] Thus, the body starts the antigen expression itself.[1] Simultaneously, the antigens are placed on the outer surface of the cells.[1] Finally, the immune system recognizes the foreign substance and starts to produce a suitable antibody for the specific antigen.[1] However, one disadvantage of using this process is that mRNA can be broken down readily.[20] Hence, it might be harder to produce enough antigens and subsequently enough antibodies to protect the body from the dreadful effects of pathogens. To reduce this disadvantage, two of the approved COVID-19 vaccines by Pfizer-BioNTech and Moderna are made utilizing self-replicating mRNAs so that the number of mRNAs is increased as was depicted in Figure 2.[13]

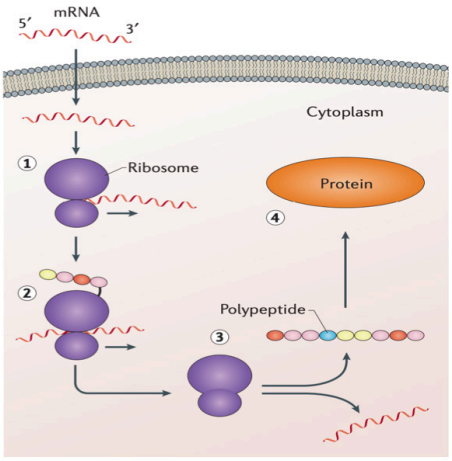


Figure 4: Working process of mRNA when it enters a cell.[19]

Efficacy of Traditional and mRNA Vaccines:

In the early usage of mRNA in research for vaccination, mostly naked mRNA was used.[21] This was creating several problems. The most important one was the instability of the mRNA which decreased the efficacy of the vaccines containing mRNA.[21] The mRNA was not able to reach the target cells safely because of the presence of nucleases which are the enzyme type that can break the mRNA down in the body.[21] Additionally, to reach the cytoplasm of the cell, mRNA needs to pass from a negatively charged phospholipid bilayer.[21] Only particles smaller than 1,000 Da (atomic mass unit) can pass through this bilayer.[21] Since mRNA is a much larger and heavier molecule, it required a carrier with it so that it could pass through the negatively charged bilayer.[21] Briefly, the deficiency of a well-prepared delivery system was the root of the problems.[21] However, this problem is no longer the case for mRNA vaccination due to the technological advancements in the mRNA delivery systems. There are several novel delivery systems divided into two big groups: viral and non-viral vector delivery systems.[18] Non-viral vector delivery systems can further be divided into two subgroups lipid and polymer delivery systems.[21] Some examples of lipid delivery systems are lipid nanoparticles and liposome complexes.[21] Some examples of polymer delivery systems are poly-amido-amine (PAA), poly-beta amino-esters (PBAEs), and polyethyleneimine (PEI).[21] Some examples of viral vector delivery systems are electroporation, polymer delivery system (polylysine), and cationic liposome.[21] Some examples of viral vector delivery systems are hybrid adenoviral vectors and retroviral vectors.[22] In this article, the main focus is on the non-viral lipid delivery system of lipid nanoparticles since the Pfizer-BioNTech, and Moderna vaccines both used this type

that mRNA can be broken down readily.[20] Hence, it might be harder to produce enough antigens and subsequently enough antibodies to protect the body from the dreadful effects of pathogens. To reduce this disadvantage, two of the approved COVID-19 vaccines by Pfizer-BioNTech and Moderna are made utilizing self-replicating mRNAs so that the number of mRNAs is increased as was depicted in Figure 2.[13]

In general, a rather popular method for solving optimization problems is Dynamic Programming (DP).[7] DP is essentially

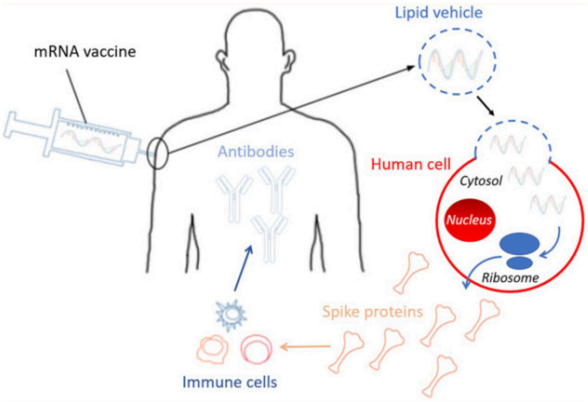


Figure 5: Fusion of lipid vehicle with the plasma membrane.[24]

a recursive solution with repeated method calls, storing the results of subproblems for the eventual solution. Compared to the aforementioned genetic algorithms, DP solutions usually use a lot of memory and have a longer running time. Even though genetic algorithms are typically harder to implement, efficiency in time and memory capacity are highly valuable assets for competitive programming which makes genetic algorithms more effective in Olympiad in Informatics.[8]

Results and Discussion

Genetic Algorithms can be considered systematic random search algorithms where the random search algorithm also considers the optimality of previous trials and evolves accordingly. Randomly initializing the individuals causes the proposed solutions to span the search space. With each generation, the solutions then attempt to converge at a global optimum. Mutations add genetic diversity to the gene pool and aim to ensure that the point the population converges is the global optimum. Knowing these it can still be inferred that a genetic algorithm may not be able to converge on the most optimal solution even with the best possible parameters as it is still a randomized and probabilistic algorithm. Although that may be the case, it could be expected that the final result of a genetic algorithm is close to the global optimum. For the set of runs that will be conducted in this paper, it is assumed that the problem is solved if the value of the solution proposed by the algorithm is within 5% of the actual solution after 2000 generations. The percentage values of how close the solution is to the actual solution are calculated using the following formula:

Another factor to consider is that genetic algorithms are intrinsically parallel compared to most algorithms which are serial. Serial algorithms only permit linear exploration that does not store every value reached by the algorithm's “infer” steps. On the contrary, genetic algorithms contain multiple offspring which allows multiple dimensional explorations. Depending on the accuracy of the offspring, some are eliminated, whereas some are pursued as promising solutions. Due to the parallelism, they can operate resembling a graph structure, evaluating multiple traceries at once. Thanks to their structure, particularly problems containing vast amounts of potential solutions are easier to solve which would take an immense amount of time with a simple search algorithm.

Identifying the sections of a problem:

First and foremost, it is essential to identify crucial sections of a problem task before analyzing any algorithms or writing code. In this problem titled IPL, a certain number of games (N) are given according to the user's input. Afterward, N number of games is expected from the user with each containing a particular fee awarded if played in the game. The problem calls for the maximum amount of money that could be earned for an individual (Nikhil) through the N games. However, there is a condition in which an individual is obligated to play at most 2 consecutive games. Herewith, a certain algorithm must be constructed which maximizes salary while taking at most 2 consecutive game conditions into account.

Furthermore, a prediction about the complexity of a problem could be made per the given test data constraints. A maximum of $2*10^5$ number of games along with 10^4 for each game can be inputted by the user. An average computer is able to make approximately $2*10^6$ calculations per second regardless of the operating system. Thus, a brute force algorithm with time complexity of $O(n^2)$ would be sufficient for a solution to pass this problem. The time complexity of $O(n^2)$ signifies the worst solution possible for a problem which usually persists in iterating through every single value until the correct answer is reached. Nonetheless, a brute force solution wouldn't pass for this problem if the user could input $2*10^8$ number of games. In a hypothetical case, a genetic algorithm would be one of the most ideal solution methods for this problem with

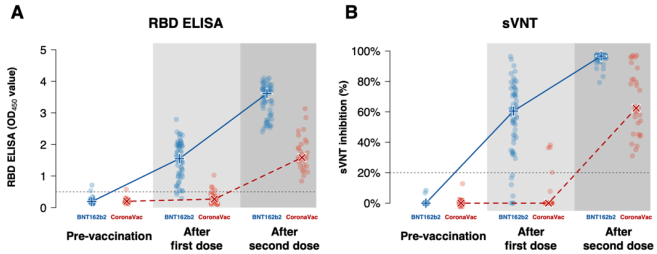


Figure 6: Comparison of BNT162b1 and CoronaVac vaccines antibody production and neutralization.[29]

the time complexity of O(gnm). Among the time complexity of O(gnm), "g" stands for the number of generations, "n" stands for the population size, whereas "m" signifies the size of individuals.[10] In brief, although in this particular case it's not imperative, a genetic algorithm solution would be a fast and viable solution to the problem among random search and brute force algorithms.

Genetic Algorithms Implementation:

The genetic algorithm is implemented via the programming language C# which is a general-purpose, object-oriented programming language developed by Microsoft. The following

DOI: 10.36838/v4i3.5

section breaks down the code into several sections - namely initialization, selection, crossover, mutation, and termination.

Before the function of the code is analyzed, its inner structure must be understood.[11] First of all, the code starts with a class named Individual, which represents one individual in the entirety of the population. Each contains a set of genes that are of template type, a double fitness value, and a chromosome length which is common throughout the entire population. Its implementation is as follows:

All of these instance variables are initialized with a constructor as each instance of the Individual class is created. Another structure within the algorithm is the Population class. The class contains an array of individuals, a generation number, a population size, a chromosome size, and various probability values for the genetic algorithm to work on. The implementation of its instance variables is as follows:

As evident from the code snippet, the population class also contains two functions: GetRandomGene() and CalculateFitness() functions. These functions are provided by the client of these classes and are called within the population's various functions. GetRandomGene() is the function that returns a random gene to initialize the members. CalculateFitness() is a function that returns a double value according to the genes of the individual; the number returned signifies the fitness of that particular gene set or individual. As cognizant of solely the meaning and implementation of these values, it is possityptype of delivery system in their SARS-CoV-2 vaccines.[23] This new lipid nanoparticle technology used by both of the approved and the most used COVID-19 vaccines not only allows the messenger ribonucleic acid to travel safely inside the body but also eases the entrance of the large and heavy mRNA molecule into the cell from a negatively charged phospholipid bilayer.[21] Whilst mRNA is entering the target cell, the encapsulating lipid fuses with the plasma membrane as described in Figure 5 by a model.[21]

Other than newly developed delivery systems, the effectiveness data of COVID-19 vaccines can be compared to show that the lipid nanoparticle delivery system has made the mRNA vaccines much more efficient than classical vaccines, specifically inactivated vaccines. In a study conducted at 99 centers in the United States where 96% of 30,420 volunteers received two doses of intramuscular mRNA-1273 injections 28 days apart, the efficacy of two doses of mRNA-1273 (Moderna) vaccine against symptomatic COVID-19 including severe diseases in persons 18 years old or older was found to be 94.1%, and no safety concerns were identified.[25] Yet, only transient local and systematic reactions were observed.[25] In the meantime, in the phase three trial of the BNT162b1 COVID-19 vaccine where 43,000 volunteers from roughly 150 clinical trial sites in several countries including the United States, Turkey, and Argentina, the overall efficacy of the vaccine was found to be 95% and the effectiveness of persons over 65 years old was recorded as 94%.[26] On the other hand, the efficiency data for the CoronaVac (Sinovac) vaccine which is a type of inactivated vaccine is lower than the ones found in mRNA vaccines when both are compared to each other. In a study in Indonesia where 1,620 healthy adults within an age range of 18 to 59 were injected with two doses of CoronaVac (Sinovac) inactivated vaccine 14 days apart, the effectiveness of the vaccine to prevent symptomatic COVID-19 was confirmed as 65.30%.[26] In a similar study conducted in Turkey, the same inactivated vaccine's efficacy against symptomatic COVID-19 was found to be 83.5%.[27] In another similar study conducted in Brazil where 43,774 people being 70 years old or more were given two shots of CoronaVac inactivated vaccine, the adjusted vaccine effectiveness against symptomatic COVID-19 was confirmed as 24.7% between days 0 and 13 meanwhile this value was found as 46.8% after the second dose.[28] In another study where the immunogenicity of mRNA vaccines and inactivated vaccines against COVID-19 based on BNT162b1 mRNA vaccine and CoronaVac inactivated vaccine were compared, the resulting graphs showed that both the amount of antibody produced by the body and the neutralization amount is higher in the BNT162b1 mRNA vaccine compared to CoronaVac inactivated vaccine as can be seen from the graph in Figure 6.[29] The relationship between antibody and antigen was measured by a method called "ELISA" as shown in graph A in Figure 6.[29] The neutralizing antibodies, on the other hand, is measured via sVNT assay.[29] The neutralizing antibodies are crucial since having more capacity for neutralization may prevent the entrance of pathogens into the cells as well as stop the germs from changing their conformational shape.[30] All in all, all the data provided shows that the efficacy rate for mRNA vaccines which are BNT162b1 and mRNA-1273 is higher than the efficacy rates for an inactivated vaccine which is in this case CoronaVac (Sinovac).

At the moment, there are not any approved live-attenuated COVID-19 vaccines since they are still in the clinical trial process which makes it harder the comparison of live-attenuated vaccines to mRNA vaccines. However, logically, some reasoning for why live-attenuated vaccines cannot perform as effectively as mRNA vaccines on COVID-19 might be done. Most importantly, since live-attenuated vaccines contain the weakened version of a virus or a bacterium, they can only attack one specific virus or bacterium type meaning that whenever there is a new pathogenic agent, a new purification and testing process is required to develop a new vaccine.[1] On the other hand, mRNA vaccines can be standardized meaning that with minimal changes in the mRNA encoding different vaccines against different disease-causing microorganisms might be produced.[1] Additionally, because of this reason, using mRNA vaccines will increase the speed of the vaccine development process which is very necessary for outbreaks such as COVID-19.[1]

Biosafety of Traditional and mRNA Vaccines:

A final aspect that should be considered while comparing the advantages and disadvantages of traditional and mRNA vaccines is their biosafety. Most of the hesitance to getting injected with an mRNA COVID-19 vaccine comes from three points: the degree of degradation of DNA of the person, instability of mRNA, and side effects of the vaccine.[21,31] To begin with, mRNA COVID-19 vaccines cannot degrade the human genome (DNA) since mRNA does not enter the nucleus of the cells where DNA is located in each person and instead remains in the cytosol which is in the cytoplasm that is found outside

of the nucleus of the cells.[21] Moreover, mRNA is a molecule that breaks down readily in the human body that's the reason why self-amplifying mRNAs are used in mRNA COVID-19 vaccines to increase the amount of mRNA in the cells which will also augment the amount of antigen and therefore antibodies.[20] Secondly, the instability of mRNA is resolved due to the advancements in the delivery systems by using appropriate carriers in COVID-19 mRNA vaccines which helps to increase the safety of the vaccines.[21] The most commonly used carriers, today, for mRNA COVID-19 vaccines are lipid nanoparticles that belong to lipids which are one of the four organic compound types.[21] Therefore, they are biodegradable meaning that they are not harmful to human cells.[20] Moreover, they fuse with the phospholipid bilayer of human cells once the mRNA encapsulated in the lipid vehicle arrives at the target cell as was shown in Figure 5.[24] Lastly, the fear of side effects of the COVID-19 vaccine. mRNA vaccines indeed have side effects such as swelling, pain, redness where the shot was administered, headache, and fever.[32] Yet, these side effects are also available from conventional vaccines including live-attenuated and inactivated vaccines. In live-attenuated vaccines, especially, people are given the weakened version of the pathogen which increases the possibility of side effects in live vaccines since the substance given by the vaccine is the closest to the actual pathogenic agent.[33] Furthermore, in live-attenuated vaccines, there is the potential for reversion to natural virulence creating a very similar effect to the actual dreadful pathogen as a result of a secondary mutation of the attenuated microorganism.[18] Moreover, because of this reason, one important limitation of live-attenuated vaccines is that they cannot be injected into people with weak immune systems since the vaccine might produce the actual disease in the body.[18] On the other hand, inactivated vaccines do not possess the possibility of turning back to natural virulence.[18] Thus, people with weakened immune systems can get these vaccines.[18] However, for this type of vaccine, a minor fault in the bacteria culturing process might result in infections in the injected people.[34] Hence, special laboratory facilities that are very clean are required.[34] A final problematic situation with inactivated vaccines is that they might require adjuvants which is an ingredient used in vaccines to boost the vaccines such as aluminum hydrates.[35] This is a disadvantage for dead vaccines because adjuvants can cause more local reactions such as swelling and pain and systematic reactions such as fever and chills.[35] This means that inactivated vaccines in which adjuvants are present have more possibility of causing side effects compared to mRNA vaccines.

Conclusion

In conclusion, several benefits of messenger ribonucleic acid (mRNA) vaccines are observed when compared to conventional vaccines comprising of live-attenuated vaccines and inactivated (dead) vaccines in manufacturing, efficiency, and biological safety. The main advantage of mRNA vaccines in production is the less requirement of time as opposed to classical vaccines which require more time due to the methods used in the discovery and development stage of the traditional vaccine development. When it comes to the efficacy of mRNA vaccines, even though previously, the instability of naked mRNA vaccines was causing a decrease in the efficiency of mRNA vaccines, the novel advancement in the technology allowed amelioration in the delivery systems which highly accelerated the augmentation of the efficacy of mRNA vaccines. Finally, the properties of mRNA as well as the technological advancement made the point of view toward mRNA vaccines more appealing. Based on these three-argument it can be concluded that mRNA vaccines provide multi-benefits over classical vaccines.

References

1. Be unique, treat individualized. https://biontech.de/covid-19-portal/mrna-vaccines (accessed Jan 30, 2022).
2. Different types of vaccines. https://carrington.edu/blog/different-types-of-vaccines-2/ (accessed Jan 30, 2022).
3. A. Patricia Wodi, MD and Valerie Morelli, ba immunology https://www.cdc.gov/vaccines/pubs/pinkbook/downloads/prinvac.pdf (accessed Jan 29, 2022).
4. Step T vaccine storage and handling. https://www.immunize.org/guide/pdfs/vacc-adults-step3.pdf (accessed Jan 29, 2022).
5. Types of vaccine. https://vk.ovg.ox.ac.uk/vk/types-of-vaccine (accessed Jan 29, 2022).
6. Content. https://vaccine-safety-training.org/inactivated-whole-cell-vaccines.html (accessed Jan 30, 2022).
7. Affairs, O. of R. Lyophilization of Parenteral (7/93). https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/inspection-guides/lyophilization-parenteral-793 (accessed Jan 29, 2022).
8. ACIP storage and handling guidelines for immunization. https://www.cdc.gov/vaccines/hcp/acip-recs/general-recs/storage.html (accessed Jan 29, 2022).
9. Office of Infectious Disease and HIV/AIDS Policy (OIDP). Vaccine types. https://www.hhs.gov/immunization/basics/types/index.html (accessed Jan 29, 2022).
10. Dolgin, E. The tangled history of mrna vaccines. https://www.nature.com/articles/d41586-021-02483-w (accessed Jan 30, 2022).
11. Pardi, N.; Hogan, M. J.; Porter, F. W.; Weissman, D. MRNA vaccines - a new era in Vaccinology. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5906799/ (accessed Jan 30, 2022).
12. What do we know about the mrna vaccine and cytokine storm? https://health-desk.org/articles/what-do-we-know-about-the-mrna-vaccine-and-cytokine-storm (accessed Jan 30, 2022).
13. Kowalzik, F.; Schreiner, D.; Jensen, C.; Teschner, D.; Gehring, S.; Zepp, F. MRNA-based vaccines. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8103517/ (accessed Jan 30, 2022).
14. J, L. I. The emergence of inactivated vaccines: Bharat Biotech. https://www.bharatbiotech.com/blog/the-emergence-of-inactivated-vaccines/ (accessed Jan 30, 2022).
15. Vaccine development, testing, and regulation. https://www.historyofvaccines.org/content/articles/vaccine-development-testing-and-regulation (accessed Jan 30, 2022).
16. voxdotcom. MRNA vaccines, explained. https://www.youtube.com/watch?v=mvA9gs5gxNY (accessed Jan 30, 2022).
17. The fastest vaccine in history. https://connect.uclahealth.org/2020/12/10/the-fastest-vaccine-in-history/ (accessed Jan 30, 2022).
18. Attenuated vaccine. https://www.sciencedirect.com/topics/immunology-and-microbiology/attenuated-vaccine (accessed Jan 30, 2022).
19. Hajj, K. A.; Whitehead, K. A. Tools for translation: Non-viral materials for therapeutic mrna delivery. https://www.nature.com/articles/natrevmats201756 (accessed Jan 30, 2022).
20. How does a mrna vaccine compare to a traditional vaccine?

DOI: 10.36838/v4i4.5

https://www.vumc.org/viiii/infographics/how-does-mrna-vaccine-compare-traditional-vaccine (accessed Jan 30, 2022).
21. Liang, Y.; Huang, L.; Liu, T. Development and delivery systems of mrna vaccines. https://www.frontiersin.org/articles/10.3389/fbioe.2021.718753/full (accessed Jan 30, 2022).
22. Nayerossadat, N.; Maedeh, T.; Ali, P. A. Viral and nonviral delivery systems for gene delivery. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3507026/ (accessed Jan 30, 2022).
23. Sanders, N. N. Low-dose single-shot covid-19 mrna vaccines lie ahead. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8134890/#:~:text=Currently%20the%20COVID%2D19%20mRNA,at%20least%20six%20times%20faser. (accessed Jan 30, 2022).
24. Anand, P.; Stahel, V. P. Review the safety of covid-19 mrna vaccines: A Review. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8087878/ (accessed Jan 30, 2022).
25. Baden, L. R.; El Sahly, H. M.; Essink, B.; Kotloff, K.; Frey, S.; Novak, R.; Diemert, D.; Spector, S. A.; Rouphael, N.; Creech, C. B.; McGettigan, J.; Khetan, S.; Segall, N.; Solis, J.; Brosz, A.; Fierro, C.; Schwartz, H.; Neuzil, K.; Corey, L.; Gilbert, P.; Janes, H.; Follmann, D.; Marovich, M.; Mascola, J.; Polakowski, L.; Ledgerwood, J.; Graham, B. S.; Bennett, H.; Pajon, R.; Knightly, C.; Leav, B.; Deng, W.; Zhou, H.; Han, S.; Ivarsson, M.; Miller, J.; Zaks, T.; COVE Study Group. Efficacy and safety of the mrna-1273 SARS-COV-2 vaccine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7787219/ (accessed Jan 30, 2022).
26. Fadlyana, E.; Rusmil, K.; Tarigan, R.; Rahmadi, A. R.; Prodjosoewojo, S.; Sofiatin, Y.; Khrisna, C. V.; Sari, R. M.; Setyaningsih, L.; Surachman, F.; Bachtiar, N. S.; Sukandar, H.; Megantara, I.; Murad, C.; Pangesti, K. N. A.; Setiawaty, V.; Sudigdoadi, S.; Hu, Y.; Gao, Q.; Kartasasmita, C. B. A phase III, observer-blind, randomized, placebo-controlled study of the efficacy, safety, and immunogenicity of SARS-COV-2 inactivated vaccine in healthy adults aged 18-59 years: An interim analysis in Indonesia. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8461222/ (accessed Jan 30, 2022).
27. CoronaVac efficacy data from Turkey. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(21)02288-1/fulltext (accessed Jan 30, 2022).
28. Ranzani, O. T.; Hitchings, M. D. T.; Dorion, M.; D'Agostini, T. L.; de Paula, R. C.; de Paula, O. F. P.; Villela, E. F. de M.; Torres, M. S. S.; de Oliveira, S. B.; Schulz, W.; Almiron, M.; Said, R.; de Oliveira, R. D.; Vieira da Silva, P.; de Araújo, W. N.; Gorinchteyn, J. C.; Andrews, J. R.; Cummings, D. A. T.; Ko, A. I.; Croda, J. Effectiveness of the CORONAVAC vaccine in older adults during a gamma variant associated epidemic of covid-19 in Brazil: Test negative case-control study. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8377801/ (accessed Jan 30, 2022).
29. Supplementary appendix - thelancet.com. https://www.thelancet.com/cms/10.1016/S2666-5247(21)00177-4/attachment/ea575ee3-b3a2-4ec9-b361-303fed53d75e/mmc1.pdf (accessed Jan 30, 2022).
30. Zoppi, L. What are neutralizing antibodies? https://www.news-medical.net/health/What-are-Neutralizing-Antibodies.aspx#:~:text=Neutralizing%20antibodies%20can%20also%20stop,the%20harmful%20effects%20of%20toxins. (accessed Jan 30, 2022).
31. Watch out for misinformation about mrna and inactivated virus COVID-19 vaccines. https://www.gov.sg/article/watch-out-for-misinformation-about-mrna-and-inactivated-virus-covid-19-vaccines (accessed Jan 30, 2022).
32. Katella, K. Comparing the COVID-19 vaccines: How are they different? https://www.yalemedicine.org/news/covid-19-vaccine-comparison (accessed Jan 30, 2022).
33. Possible side effects from vaccines. https://www.cdc.gov/vaccines/vac-gen/side-effects.htm (accessed Jan 30, 2022).
34. The different types of covid-19 vaccines. https://www.who.int/news-room/feature-stories/detail/the-race-for-a-covid-19-vaccine-explained (accessed Jan 30, 2022).
35. Adjuvants and vaccines. https://www.cdc.gov/vaccinesafety/concerns/adjuvants.html (accessed Jan 30, 2022).

Authors

Deniz Cenikli is a sophomore student at Uskudar American Academy in Turkey. She is currently in the IB diploma program. She immersed herself in biology, chemistry, and mathematics. As a student willing to study medicine, she is eager to enhance her scientific knowledge.

IJHSR
ijhighschoolresearch.org

■ RESEARCH ARTICLE

Hierarchical Optimal Path Planning (HOPP) for Robotic Apple Harvesting

David W Liu
Athens Academy, 1281 Spartan Ln, Athens, GA, 30606, USA; david.weizhong.liu@gmail.com

ABSTRACT: Apples are among the most consumed fruits in the United States. Currently, almost all apples destined for the fresh market are picked by the human hand. Due to the shortage of seasonal manual labor and rising costs in manual apple picking, robotic apple harvesting has been explored for years. However, a challenge in developing and deploying an apple harvesting robotic system is how to deal with the unstructured apple tree environment. This work aims to demonstrate that structured 3D model representation of apple trees can significantly enable and facilitate robotic apple harvesting, particularly for optimal 3D path planning. Accordingly, a hierarchical optimal path planning (HOPP) algorithm is designed to significantly reduce or minimize the time cost during robotic apple harvesting in a 3D environment. The core idea of this HOPP algorithm is applying distance-constrained k-means clustering to group apples into 3D harvesting zones first, after which an optimal 3D robotic harvesting path is derived via the Traveling Salesman Problem (TSP) formulation and solution. Within each 3D apple harvesting zone, a second-stage optimal path planning is conducted by the TSP method on individual apples. Experiments showed that the proposed HOPP algorithm is promising.

KEYWORDS: Robotics; Agriculture; Apple Harvesting; Path Planning.

■ Introduction

Apples are America's most consumed and favorite fruit and are either eaten raw or consumed in other ways. The American apple industry, with 5,000+ producers, brings billions of dollars to the economy each year and contributes nutritious fruit/food to consumers around the world. Remarkably, the most labor-intensive task in apple production is harvesting, and nearly all apples destined for the fresh market in America are picked by the human hand. Apple growers reported that harvesting labor accounts for approximately a third of their annual variable costs - as much as pruning and thinning combined.[1] In addition, manual apple harvesting is a time-sensitive operation where variable weather patterns generate uncertainty during employment planning and is costly.[1] For example, the threat of an early fall frost could cause a short-term surge in the demand for apple pickers. Also, picking fresh market apples is both physically strenuous and highly repetitive. Apple picking exposes workers to fall hazards as well as ergonomic injuries through heavy lifting and repetitive hand actions. Therefore, to deal with the shortage of seasonal manual labor, rising costs, and risks to workers in manual apple picking, robotic apple harvesting has been explored for decades.[2]

Robotics researchers have been actively working on the development of apple harvesting robots since the 1980s.[2] In this field, visual fruit harvesting robots are common as reviewed recently.[3] Typically, a fruit harvesting robot uses visual sensing, e.g., via 2D cameras, stereo vision systems, laser active vision systems, or multi-spectral imaging technologies, to perceive and learn fruit information such as location, shape, size, color, etc. The main technical tasks of those visual sensing systems include camera calibration, fruit target recognition and positioning, background recognition, 3D fruit reconstruction, and robotic path and trajectory planning. Those visual sensing systems also use visual servocontrol picking mechanisms to perform real fruit harvesting. However, a notable missing piece of those existing visual sensing systems for fruit harvesting is that they are not designed for global 3D mapping or reconstruction of all fruits on the entire tree, e.g., all apples on a tree. Instead, existing visual sensing techniques in robotic fruit harvesting are targeting individual or a few fruits in the local context. The lack of global information of distributions or localizations of all apples on a tree hampers the possibility of global optimal 3D path planning for actual apple harvesting. In response, this study contributes a unique investigation of how much gain can be achieved for optimal 3D path planning in robotic apple harvesting if 3D models of apples and their trees are available, which can stimulate and motivate future innovation in developing 3D reconstruction and mapping approaches for robotic apple harvesting and robotic fruit harvesting in general.

This work is also motivated by a fact that unstructured apple tree environment is a major challenge in developing and deploying apple harvesting robotic systems in real world scenarios, as pointed out and emphasized by literature papers.[1-3] By leveraging an existing realistic 3D model of an apple tree and its hundreds of apples, this work will demonstrate that 3D representation of apples can significantly enable and facilitate robotic apple harvesting, particularly for optimal 3D path planning. Along this direction, a hierarchical optimal path planning (HOPP) algorithm is designed to significantly reduce or minimize the time cost during robotic apple harvesting in a 3D environment. Specifically, the 3D HOPP algorithm is composed of two stages: 1) applying distance-constrained k-means clustering4 to group apples into harvesting zones first, based on which an optimal

DOI: 10.36838/v4i4.6

robotic harvesting path is derived via the Traveling Salesman Problem (TSP) formulation and solution;[5] 2) within each apple harvesting zone, a second-stage optimal path planning is conducted by the TSP solution on individual apples. The above two stages of optimization form the HOPP algorithm in order to significantly reduce or minimize time cost in robotic apple harvesting. This work differentiates itself from other existing studies by contributing a new concept that 3D representation of apples can significantly enable and facilitate robotic apple harvesting and by contributing a novel HOPP algorithm.

Methods

3D environment setup:

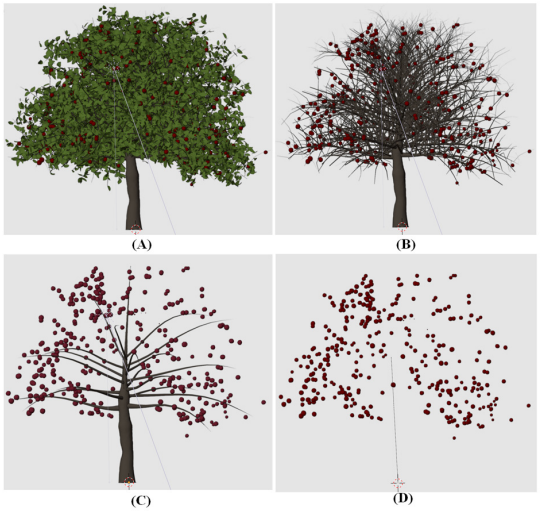


Figure 1: Illustration of the 3D model of an apple tree and its apples. (A) Full 3D model of the apple tree. (B) Leaves are removed from the tree. (C) Small branches are removed further. (D) Trunk is removed further and only apples are kept.

This work uses a commercially purchased realistic 3D model of apple trees (from TurboSquid website) that includes 349 apples on it (represented in obj file format). Specifically, the apple tree's 3D model includes 640,768 vertices and 458,022 faces. Figure 1 illustrates several views of the apple tree and its apples. All of the HOPP algorithm developments and evaluations in this work are based on the 3D model in Figure 1. According to the 3D rendering of apples, it is observed that spatial distributions of apples are quite irregular, e.g., some regions have higher apple densities, while other regions have lower apple densities. Therefore, a straightforward intuition is that such irregular spatial distribution of apples should be leveraged for optimal 3D path planning in robotic apple harvesting, that is, the apple harvesting robot should give higher priority to those regions with higher apple densities and plan its 3D path accordingly.

The HOPP algorithm:

Time cost is a key concern in robotic apple harvesting, e.g., the time costs of moving robotic shoulders and arms. In this work, I divided the total time cost of harvesting all apples into two stages: 1) moving the robotic shoulder across harvesting zones and 2) moving robotic arms to pick individual apples within each harvesting zone. Specifically, assume that the total number of robotic shoulder movements is $k-1$, that is, there are k harvesting zones, and each movement's average time cost is tshoulder. Then the total time cost of robotic shoulder movements is $(k-1)^*t_{shoulder}$. Within the workspace of robotic arms (e.g., simplified as a 3D sphere centered at the joint of robotic shoulder and robotic arms) in each harvesting zone, the average time cost for picking an individual apple when only moving the robotic arms is represented by tarm. Then, to pick all the apples (n) on a tree, the time needed for individual apple harvesting when moving the robotic arms is n^*t_{arm}. Therefore, the total time cost of harvesting all apples, including both robotic arm and shoulder movements, will be $(k-1)^*t_{shoulder}+ n^*t_{arm}$, which is the objective function of the HOPP algorithm to significantly reduce or minimize.

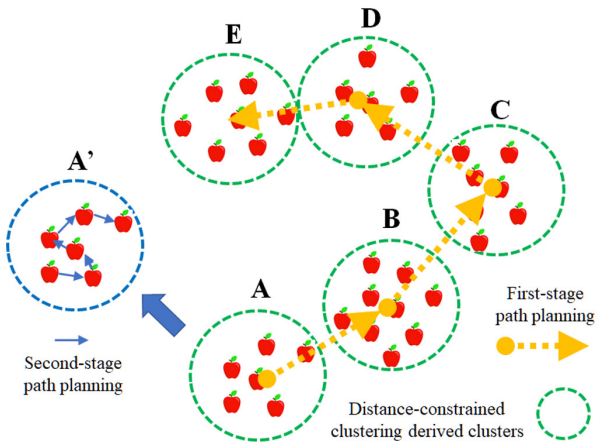


Figure 2: Overall idea of the HOPP algorithm. Here A-E represent five harvesting zones obtained by the distance-constrained *k*-means clustering. The path represented by orange arrows shows the first-stage optimal path across harvesting zones. The apples in A' are replicated from A. The blue arrows in A' show the second-stage optimal path across individual apples within a harvesting zone.

The overall idea of the 3D HOPP algorithm is simplified in a 2D space and illustrated in Figure 2. Specifically, in the first stage, all the apples will go through a distance-constrained k-means clustering so that these apples will be geographically grouped into neighboring harvesting zones (represented by the dashed green circles in Figure 2). Here, the harvesting robot arm's workspace is simplified by a circle. In the 3D case, the robotic arms can reach anywhere in the 3D spherical volume in the harvesting zone without the need of moving the robotic shoulder. Then, the optimal robotic path connecting those harvesting zones is formulated as the TSP and is solved using heuristic greedy search solutions, as illustrated by the dashed orange arrows (A->B->C->D->E) in Figure 2. Similarly, in the 3D case, the optimal path is obtained by connecting those 3D spheres via the TSP solution. Intuitively, the step of distance-constrained k-means clustering can minimize the total number of apple harvesting zones. That is, given the fixed diameter of robotic arms' spheric workspace, the distance-constrained *k-means* clustering can ensure that each grouped harvesting zone will include the largest possible number of apples, thus the total number of harvesting zones will be minimized. Also, the TSP solution to the cross-harvesting zone path can further minimize the average time cost of robotic shoulder movement. As a result, the total number

of robotic shoulder movements $k-1$ and each shoulder movement's average time cost is $t_{shoulder}$ are both minimized. Therefore, the total time cost of robotic shoulder movements $(k-1)^*t_{shoulder}$ is minimized.

In the second stage, within each apple harvesting zone, an optimal path planning is further conducted on individual apples by the TSP formulation and solution again, as illustrated in the dashed blue circle and blue arrows in A' on the left side of Figure 2. Therefore, the average time cost (t_{arm}) for picking an individual apple within each harvesting zone is minimized, and the total time needed for harvesting all individual apples (n^*t_{arm}) when moving robotic arms is minimized. Thus, through the HOPP algorithm, the total time cost of harvesting all apples ($(k-1)^*t_{shoulder} + n^*t_{arm}$), including time costs for both robotic arm and shoulder movements, is minimized.

Distance-constrained k-means clustering for grouping apples into harvesting zones:

Data clustering is a commonly employed methodology to group or partition a set of observational data (e.g., apples in this work) into collections with small within-group distances (e.g., within each apple harvesting zone) and big cross-group distances (e.g., across apple harvesting zones). Among various data clustering methodologies, one of the most popular methods is the k–means clustering algorithm.[6] Briefly, the k-means clustering algorithm utilizes iterative refinement to generate a data clustering result after taking the number of clusters k and the dataset (e.g., apples' spatial coordinates in this work) as the input. The k-means clustering algorithm starts with an initial estimate for the k centroids, which can either be randomly created or randomly selected from the dataset. The algorithm then iterates between two steps of data assignment and centroid update. More specifically, in the data assignment step, each centroid defines one cluster, and each data point is assigned to its nearest centroid based on the metric of Euclidean distance, e.g., distance between apples' spatial coordinates in this work. In the centroid update step, the cluster centroids are recomputed and updated by averaging the coordinates of all data points (e.g., apples' spatial coordinates here) that are assigned to that centroid's cluster. The k-means clustering algorithm iterates between the above two steps until a stopping criterion is met, i.e., no data points change clusters, the sum of the distances is minimized, or a maximum number of iterations is performed.

However, in this work's context, directly applying the k-mean clustering algorithm to the apples' data points would not work well due to the following reasons. First, it will be challenging to determine the right selection of cluster number k in our application scenario if the k-means clustering algorithm is directly applied. Second, it is hard to satisfy that the derived clusters will form valid apple harvesting zones, that is, all apples within a harvesting zone can be reached by the robotic arms, meaning that the distance between any apple and the centroid of the harvesting zone is shorter than the maximum length of the robotic arms. Therefore, a modified variant of k-mean clustering is adopted in this work's context, that is, the distance-constrained k-mean clustering.[4]

Mathematically, the following problem definition of distance-constrained k-mean clustering is adopted. Consider a collection of n apples' 3D coordinate data, $x1,\ldots,xn$, with $xi \in R^3$, I aim to select a number of k groups or clusters (harvesting zones), $C1,\ldots,Ck$. It typically holds that $k \ll n$. Each cluster Cj is assigned with a cluster centroid $cj \in R^3$, which represents the center that apple cluster. An apple $xi \in R^3$ is said to belong to cluster Cj if its distance from the centroid ch, with $h \neq j$, of every other cluster Ch is larger than its distance from Cj. If xi belongs to Cj, I set a binary assignment variable $ri, j \in B$ to 1, and to 0 otherwise. The solution of the apple clustering problem with distance constraint involves the computation of the optimal choice of cluster centroids cj, for $j=1,\ldots,k$, that minimizes the total of the distance of every apple xi from the cluster it belongs to. More specifically, the problem is defined as follows.

I aim to find a choice of k, of the cluster centroids, $cj \in R^3$, and measurement data assignments, $ri, j \in B$, that minimizes the objective function E:

$$E = \sum_{i=1}^{n} \sum_{j=1}^{k} ri,j * \|xi - cj\| \quad (1)$$

Subject to the following three constraints:

$$\text{I: } \sum_{j=1}^{k} rij = 1, \ \forall i = 1, \ldots, n \quad (2)$$

$$\text{II: } \|xi - cj\| \leq \beta/2, \ \forall i = 1, \ldots, n; \ \forall j = 1, \ldots, k \quad (3)$$

$$\text{III: } rij \in \{0,1\}, \ \forall i = 1, \ldots, n; \ \forall j = 1, \ldots, k \quad (4)$$

The first constraint (I) along with the third one (III) implies that each apple is assigned exactly to one cluster (harvesting zone); the constraints (II) imply that β is the diameter of the harvesting zone and the distance between any apple in this harvesting zone and its centroid is less than or equal to its radius; finally, constraints (III) imply that rij are binary decision variables.

The definitions in Equations (1)-(4) fit our apple harvesting zone definition (illustrated in Figure 2) very well. Notably, in the literature, the constraint (II) in Equation (3) has not been considered, as far as I know. Including such a constraint, however, is quite useful in this work, since it will ensure the HOPP algorithm's ability to define valid apple harvesting zones within a predefined spherical robotic arm workspace. Meanwhile, the objective function in Equation (1) ensures that the total number of apple harvesting zones is minimized. To find a plausible solution to the problem defined in Equations (1)-(4), the method that used a Hegelsmann-Krause (HK) model was adopted in this work to find an appropriate k first before the distance-constrained k-means clustering.[4,7] The core idea of the HK model is that data points with completely different attributes (e.g., geometric distances here) do not influence each other, while some sort of mediation occurs among data points whose distances are close enough.[7] It has been justified and demonstrated that the HK model is well suited for the distance-constrained k-means cluster.[4]

Traveling salesman problem solution and hierarchical optimal 3D path planning :

Once the apples have been clustered into harvesting zones via the methods in Equations (1)-(4), the spatial relationships among neighboring harvesting zones can be represented by a

DOI: 10.36838/v4i4.6

graph, e.g., spatially adjacent harvesting zones are linked by an edge, as illustrated in Figure 2. Then, the optimal 3D path planning at the first stage of harvesting zone is formulated as a TSP and solved using existing solution.[5] Theoretically, TSP is an NP-hard problem in combinatorial optimization and a heuristic greedy search algorithm is typically employed to obtain a near optimal result. Notably, the TSP formulation and solution has been utilized in many path planning problems in agriculture robotics and automation. Here, a heuristic greedy search implementation of TSP problem was adopted to find the optimal path given the clustered apple harvesting zones and their neighborhood graph, as illustrated in Figure 2. Thus, both the total number of robotic shoulder movements k and each shoulder movement's average time cost is $t_{shoulder}$ are minimized via the distance-constrained k-means clustering of apples and the TSP solution of optimal path planning at the harvesting zone level, respectively. Through this first stage of optimal 3D path planning, the total time cost of robotic shoulder movements $(k-1)*t_{shoulder}$ is thus minimized.

Similarly, in the second stage of path planning, an optimal path search is further conducted on individual apples by the TSP formulation and solution again, as illustrated by the apples in A' of Figure 2. As a result, the average time cost t_{arm}) for picking an individual apple within each harvesting zone is minimized, and thus the total time cost for harvesting all individual apples ($n*t_{arm}$) is minimized. Finally, the two-stage HOPP algorithm is summarized in Figure 3.

Input: n apples, represented by xi $(i=1,...,n)$

Stage 1: distance-constrained k-means clustering of xi $(i=1,...,n)$ into k apple harvesting zones.

The Hegelsmann - Krause (HK) model is used to determine the best k clusters.

The robot arm workspace's sphere diameter β is used to constrain the size of each harvesting zone (Equation (2)).

The distance-constrained k-means clustering algorithm is used to group all apples into k harvesting zones (Equations (1)-(4)).

Use TSP formulation and heuristic search solution to find the 3D optimal path across k harvesting zones.

This stage achieves the minimization of the total time cost of robotic shoulder movements ($(k-1)*t_{shoulder}$).

Stage 2: Use TSP formulation and solution to find the 3D optimal path within each harvesting zone.

The stage achieves the minimization of individual apple harvesting ($n*t_{arm}$).

Output: hierarchical optimal 3D path across all harvesting zones and individual apples.

Figure 3: Summary of the two-stage HOPP algorithm for robotic apple harvesting. Details of each stage or step have been provided in the above sub-sections.

Results and Discussion

Experiments:

A series of experiments were designed and conducted to demonstrate the effectiveness and performance of the proposed HOPP algorithm summarized in Figure 3 and detailed in the above sections. The 3D models of apples in Figure 1 were used in these experiments.

As explained in Equation (3) and summarized in Figure 3, the parameter β describes the diameter of apple harvesting robot arms' spherical workspace and it is a key factor that decides the output of the distance-constrained k-means clustering algorithm (Equations (1)-(4)). By alternating different β values, various apple harvesting zone clustering results were obtained by the HOPP algorithm. For quantitative comparison, a traditional path planning method (named Method 2 here) was considered as follows. It randomly selects apples to start with the harvesting process and uses greedy search in a local neighborhood for path planning. That is, Method 2 always goes from the current apple to the next neighbor with the shortest distance. Again, the average time cost for picking an individual apple within each harvesting zone is represented by t_{arm}, and each shoulder movement's average time cost is $t_{shoulder}$. Then, I compare the total time costs for both path planning methods (HOPP and Method 2), that is, $(k-1)*t_{shoulder} + n*t_{arm}$. Here, it is assumed that the time cost $t_{shoulder}$ is proportional to the distance between two harvesting zones (either minimized by the TSP optimization in HOPP or not in Method 2), and that the time cost t_{arm} is proportional to the distance between two individual apples (either minimized by the TSP optimization by HOPP or by greedy search in Method 2). It turned out that the time costs t_{arm} and $n*t_{arm}$ are quite close for HOPP and Method 2 as both methods employed a greedy search process among neighboring individual apples. Therefore, the following experimental comparisons focus on the time costs of robotic shoulder movements $(k-1)*t_{shoulder}$.

When β=1.75 (based on the distance unit in the 3D model of apple tree in Figure 1), 16 apple harvesting zones were obtained by the HOPP method, as shown by the large color spheres in Figure 4A. The derived optimal 3D optimal path is represented and visualized by the white lines in Figure 4A and by the corresponding blue arrows Figure 4B. It is apparent in Figure 4A that each clustered harvesting zone exhibits high density of apples, meaning that each robotic shoulder movement can harvest as many apples as it can and therefore the time cost of robotic shoulder movement ($(k-1)*t_{shoulder}$) is minimized. Also, the optimal 3D path at the level of individual apples obtained by the TSP solution is represented by the white lines and shown in Figure 4C. In comparison, the path planning results by Method 2 are shown in Figures 4D-4F in a similar way. It is interesting that the planned path by Method 2 takes 40 robotic shoulder movements (for one random run) from one harvesting zone to another. That is, without taking advantage of global distribution of all apples and without the distance-constrained k-means clustering, the planned path by Method 2 takes 2.67 times the robotic shoulder movements (40 steps) to harvest all apples when compared to the HOPP method (15 steps). This result apparently suggests the effectiveness of the proposed HOPP method and the importance of taking the advantage of the global information of spatial distributions of all apples on trees for more efficient apple harvesting.

Notably, Method 2 randomly selected the start apple to harvest and thus its total steps of robotic shoulder movements is random. I repeated Method 2 30 times and the statistical results are shown in Figure 5. On average, the planned path by Method 2 takes 2.85 times the robotic shoulder movements (42.73 steps) to harvest all apples than the HOPP method (15 steps), as shown in Figure 5A. Also, the average path distances

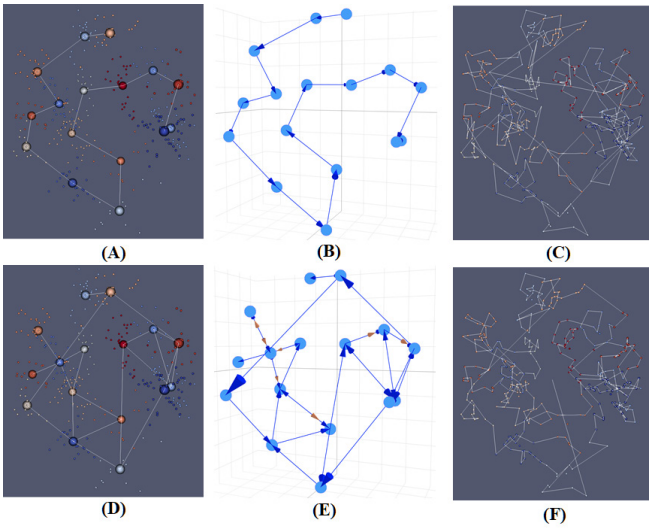


Figure 4: (A)-(C): visualization of result by HOPP. (A) The small spheres represent apples, and each color represents one harvesting zone. The large spheres are the centers of clustered harvesting zones, and their colors correspond to the apple clusters. The white lines are the path across harvesting zones. (B) The harvesting zone path in (A) is represented by blue arrows with directions. (C) Harvesting path at individual apple level. Each while line represents one step. The apple colors represent the clusters in (A). (D)-(F): The visualization of results by Method 2. In (E), for some pairs of harvesting zones, there are multiple path steps, and they are represented by multiple color arrows (blue and brown).

across harvesting zones by Method 2 (111.80) are 3.06 times more than the HOPP method (36.49), as shown in Figure 5B. Therefore, in terms of $(k-1)*t_{shoulder}$, Method 2 takes 8.72 times the harvesting time by the HOPP method.

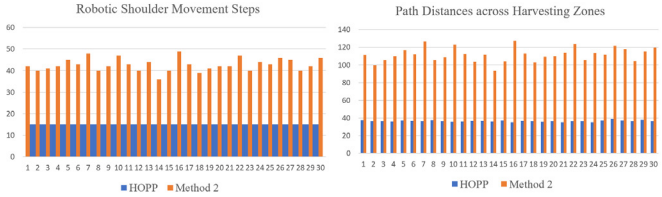


Figure 5: Comparison of HOPP and Method 2. Both methods were run for 30 times. (A) The total number of robotic shoulder movement steps in all 30 runs. The average steps for HOPP and Method 2 are 15 and 42.73, respectively. (B) Path distances across harvesting zones in 30 runs. It is noted that the initial position of HOPP slightly impacts its result. The average distances for HOPP and Method 2 are 36.49 and 111.80, respectively.

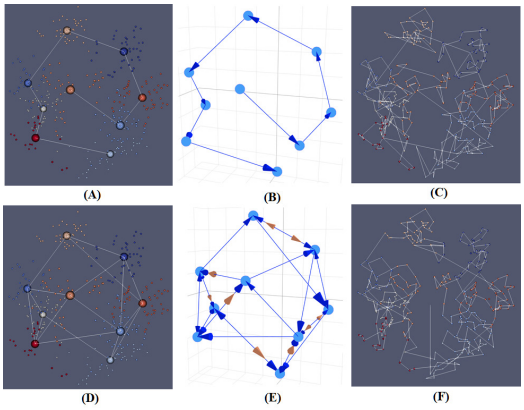


Figure 6: (A)-(C): visualization for HOPP. (A) The small spheres represent apples, and each color represents one harvesting zone. The large spheres are the centers of clustered harvesting zones, and their colors correspond to the apple clusters. The white lines are the path across harvesting zones. (B) The harvesting zone path in (A) is represented by blue arrows with directions. (C) Harvesting path at individual apple level. Each while line represents one step. The apple colors represent the clusters in (A). (D)-(F): The visualization results for Method 2. In (E), for some pairs of harvesting zones, there are multiple path steps, and they are represented by multiple color arrows (blue and brown).

When β=2.2, 9 apple harvesting zones were obtained by the HOPP method (Figures 6A-6B), and the optimal 3D optimal path is shown in Figure 6C. The path planning for Method 2 is shown in Figures 6D-6F. It is clear in Figure 6 that the proposed HOPP method takes significantly less time than Method 2. Again, Method 2 randomly selected the start apple to harvest and thus its total steps of robotic shoulder movements is random. I repeated the Method 2 for 30 times and the statistical results are shown in Figure 7. On average, the planned path by Method 2 takes 4.23 times the robotic shoulder movements (33.90 steps) to harvest all apples than that of the HOPP method (8 steps), as shown in Figure 7A. Also, the average path distances across harvesting zones by Method 2 (111.93) are 4.76 times more than those by the HOPP method (23.51), as shown in Figure 7B. Therefore, in terms of $(k-1)*t$-shoulder, Method 2 takes 20.13 times more harvesting time than the HOPP method.

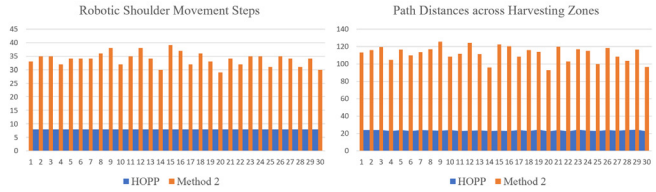


Figure 7: Comparison of HOPP and Method 2. Both methods were run for 30 times. (A) The total number of robotic shoulder movement steps in all 30 runs. The average steps for HOPP and Method 2 are 8 and 33.90, respectively. (B) Path distances across harvesting zones in 30 runs. The average distances for HOPP and Method 2 are 23.51 and 111.93, respectively.

Reconstruction of 3D apple tree model:

Following the direction of 3D path planning for robotic apple harvesting and inspired by existing efforts of 3D reconstruction from image sequences,[8] an initial effort was made to reconstruct 3D apple tree model in a lab environment. As shown in Figure 8, 63 photos (Figures 8A-8B) were taken around a laboratory apple tree by an iPhone from various angles and locations, and these photos were then used as the input for 3D reconstruction via the Meshroom software toolkit.[9] The Live Reconstruction pane in Meshroom interface provides the visualization of each step in the 3D reconstruction, such as camera initialization and calibration, feature extraction, and structure from motion. Notably, the Meshroom toolkit utilizes a camera sensors database to determine the camera's internal parameters and group them together. Therefore, Meshroom can infer the rigid scene structure (3D points) with the pose (position and orientation) and internal calibration of all cameras (iPhone camera's different poses), as shown in Figure 8C. By using the calibrated cameras and the structure-from-motion step, Meshroom can generate a dense geometric surface of the 3D reconstruction of the apple tree, as shown in Figures 8C-8F. Although the 3D model's quality can be improved in the future, the experiment in Figure

DOI: 10.36838/v4i4.6

8 demonstrated the promise of reconstructing 3D models of apple trees using computer vision and computer graphics techniques. Given the rapid advancements of technologies, such as unmanned aerial vehicles (UAV) LiDAR or UAV camera data acquisition and 3D reconstruction,[10,11] it is predicted that robotic apple harvesting guided by 3D path planning will become practical in the future

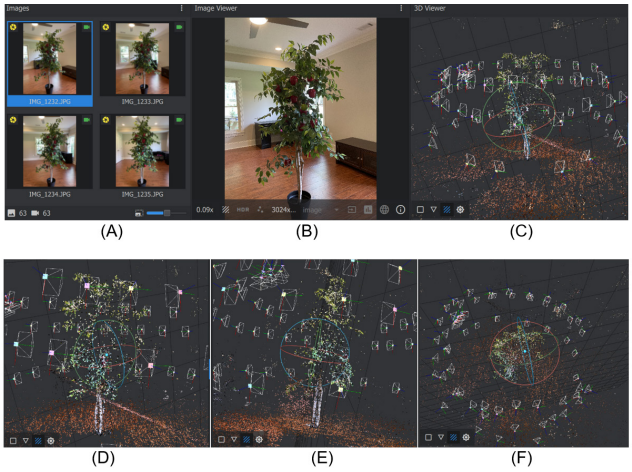


Figure 8: Reconstruction of 3D apple tree model. (A) Visualization of 63 images captured by iPhone camera from 63 poses. (B) Enlarged view of the first image in (A). (C) Reconstruction of the 3D apple tree model. (D)-(F): additional views of the 3D model in (C).

Path planning for fruit harvesting robotics:

From a general perspective of robotic path planning, the HOPP algorithm proposed in this paper belongs to off-line path planning,[12] that is, the apple harvesting robot has prior access to complete information about the 3D environment of apples and their tree. Meanwhile, this HOPP algorithm can be applied online during harvesting if more information (e.g., apple localization and 3D reconstruction) becomes available. Also, the HOPP algorithm belongs to the category of point-to-point path planning, that is, the goal consists of determining a path from a starting point to a destination point by optimizing key parameters such as time and distance.[12] The HOPP algorithm is akin to the general category of cell decomposition method in the robotic path planning domain, which decomposes the free space into small regions called cells (e.g., apple harvesting zones in this work) and then searches for an optimal path in the cell graph using algorithms such as A*, Dijkstra, or TSP.[12] The HOPP algorithm also shares similarities with the category of coverage path planning (CPP) algorithms, in that CPP is defined as the task of determining a path that passes overall points of an area or volume (e.g., apple harvesting zones here) while avoiding obstacles.[13] The HOPP algorithm and CPP share such common characteristics or requirements as the robot must cover the whole area (all apples), the robot should cover the entire region without overlapping, the robot should avoid all obstacles, the robot should use simple and smooth motion trajectories, and an optimal path is desired under considered criteria such as time and distance. However, it is not always possible to possess all these requirements or characteristics in complex agriculture environments like apple trees or orchards, and thus priority consideration is typically adopted. For instance, in its current form, the HOPP algorithm does not consider obstacle avoidance and does not plan robotic shoulder/arms motion trajectories.

There is much room to improve the HOPP algorithm in the future, such as considering obstacle avoidance, considering other optimization objectives like smoothness of robotic arms/shoulder movement, considering different time costs of robotic arm/shoulder movements, considering multiple simultaneous harvesting robots, and 3D path planning for harvesting many apple trees in the orchard. With these more complex considerations, the HOPP algorithm can be extended in various ways. For instance, the distance-constrained *k*-means clustering can be extended by adding additional obstacle avoidance into the optimization constraints, and the algorithm itself should be scaled to much larger number of apples like those in apple orchard. Also, the TSP formation and solution for finding the optimal path can be extended into a multiple TSP (mTSP) formation if there are several apple harvesting robots simultaneously participating in the apple picking.

It is known that the k-means clustering algorithm has a time complexity of $O(n^2)$, where n is the input data size. This quadratic complexity debars the algorithm from being effectively used in large-scale applications. Researchers have explored lower-complexity implementation of the *k*-means clustering algorithm, such as the $O(n)$ complexity (linear order) counterpart of the *k*-means.[14] In the future, such lower-complexity implementation could be employed into the HOPP algorithm for apple harvesting zone clustering via distance-constrained k-means clustering, which is the most time-consuming part of the 3D path planning algorithm. Also, such lower-complexity heuristic search algorithms can be explored for the TSP and mTSP solutions once the harvesting zones are mapped. These lower-complexity implementation of the 3D path planning algorithm is quite important for online path planning in large-scale apple harvesting in real-world orchards. Finally, this work does not provide a theoretical analysis of the algorithmic aspect of the HOPP method, which should be investigated in the future.

■ Conclusion

The world's population is projected to reach nine billion people by the year 2050, which suggests that agricultural productivity must increase substantially and sustainably. The automatization of agricultural tasks, including fruit harvesting, is an essential step to deal with population growth. Various types of agricultural robots have been explored in the past few decades, however, how to deal with the unstructured agricultural environment is one of the hardest challenges. This work proposes a conceptually new HOPP algorithm for robotic apple harvesting by decomposing the challenging task into two-stage hierarchical apple harvesting zone clustering and optimal 3D path planning, and experimental results have demonstrated that the HOPP algorithm can gain significant benefits, e.g., minimizing apple harvesting time cost. In addition to the HOPP algorithm, another core contribution of this work is that it is demonstrated that the global information of spatial distributions of all apples should be leveraged for better apple harvesting. Although this proposed HOPP algorithm

leverages the availability of 3D models of apples at current stage, it is expected that reconstructing 3D structures of apples and their trees from UAVs equipped with camera, LiDAR or multispectral imaging sensors and enabled with advanced AI tools will become practical in the future. Also, despite that this work simplifies the apple tree environment and harvesting execution, the HOPP algorithm itself is scalable and extendable by including a variety of other considerations, such as obstacle avoidance and multiple harvesting robots, in the future.

Acknowledgements

I would like to thank Professor Charlie Li, Dr. Javier Rodriguez and Daniel Petti of Bio-Sensing and Instrumentation Lab, Phenomics and Plant Robotics Center, College of Engineering, University of Georgia, for their guidance and help when I performed an internship in their lab in Summer 2021 and Fall 2021. I would like to thank Zhengliang Liu and Jiaxing Gao for their guidance and help on data visualization and Python programming. I would like to thank my teacher Mike Callinan at Athens Academy who has been teaching me robotics courses and guiding me on robotics projects for over four years. Finally, I would like to thank my father for his guidance and support during my research.

References

1. Silwal, A.; Davidson, J. R.; Karkee, M; Mo, C.; Zhang, Q.; Lewis, K. Design, integration, and field evaluation of a robotic apple harvester, *J Field Robotics*. 2017, 34:1140-1159.
2. Jia, W.; Zhang, Y.; Lian, J.; Zheng, Y.; Zhao, D.; Li, C.; Apple harvesting robot under information technology: A review. *International Journal of Advanced Robotic Systems*. 2020, 17(3):1-16.
3. Tang, Y.; Chen, M.; Wang, C.; Luo, L.; Li, J.; Lian, G.; Zou, X. Recognition and Localization Methods for Vision-Based Fruit Picking Robots: *A Review, Front Plant Sci*. 2020, 11:510.
4. Oliva, G.; Manna, D. L.; Fagiolini, A.; Setola, R. Distance-constrained data clustering by combined k-means algorithms and opinion dynamics filters. *The 22nd Mediterranean Conference on Control and Automation*. 2014, pp. 612-619.
5. Held, M.; Karp, R.M. The Traveling Salesman Problem and Minimum Spanning Trees. *Operations Research*. 1970, 18(6):1138-1162.
6. Hartigan, J. A.; Wong, M. A. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society, Series C*. 1979, 28(1):100-108..
7. Hegselmann, R.; Krause, U. Opinion dynamics and bounded confidence models, analysis, and simulation, *Journal of Artificial Societies and Social Simulation*. 2002, 5(3).
8. Qu, Y.; Huang, J.; Zhang, X.; Rapid 3D Reconstruction for Image Sequence Acquired from UAV Camera, *Sensors*. 2018, 18(1):225.
9. Kneip, L.; Scaramuzza, D.; Siegwart, R. A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2011, 2969-2976.
10. Hadas, E.; Jozkow, G.; Walicka, A.; Borkowski, A. Apple orchard inventory with a LiDAR equipped unmanned aerial system, *International Journal of Applied Earth Observation and Geoinformation*. 2019, 82:101911.
11. Chen. Y.; Zhang, B.; Zhou, J.; Wang, K. Real-time 3D unstructured environment reconstruction utilizing VR and Kinect-based immersive teleoperation for agricultural field robots, *Computers and Electronics in Agriculture*. 2020, 175:105579.
12. Mac, T. T.; Copot C.; Tran, D. T.; Keyser, R. D. Heuristic approaches in robot path planning: A survey, *Robotics and Autonomous Systems*. 2016, 86:13–28.
13. Galceran, E.; Carreras, M. A survey on coverage path planning for robotics, *Robotics and Autonomous systems*. 2013, 61(12):1258-1276.
14. Pakhira, M. K. A Linear Time-Complexity k-Means Algorithm Using Cluster Shifting, International Conference on Computational Intelligence and Communication Networks. 2014, pp. 1047-1051.

Author

David W Liu is a 10th grader in Athens Academy, Athens, GA. He is interested in robotics, computer programming, and mathematics. He has been taking robotics courses in Athens Academy since 6th grade and has been actively involved in multiple robotics activities such as FLL and FTC challenges since 6th grade.

DOI: 10.36838/v4i4.6

IJHSR
ijhighschoolresearch.org
■ RESEARCH ARTICLE

Why You Can't Remember Where Your Keys Are: The Effects of Sleep Deprivation on Visual Short-Term Memory

Dila Ekrem
St. George Austrian High School, Bereketzade, Kart Çınar Sk. No:2, 34420 Beyoğlu/İstanbul, Turkey; dilaekrem@gmail.com

ABSTRACT: This study examines the effects of sleep deprivation on young adults' short-term memory. High school students between ages 16-18 from Turkey and the United Kingdom (n= 106) were asked to fill out a survey, play a short-term memory game and watch a video that showed 20 photos to them for 20 seconds. I predicted that the young adults who slept less than six hours the other day would have worse short-term memory, take more time to complete the short-term memory game, and would not be able to recall the images as well as the ones who slept more than six hours. In addition to this, I predicted that the young adults who slept less would be prone to remember the negative emotions triggering images better than other images and created a memory test to examine it. Recent findings demonstrate that sleep deprivation is linked with short-term memory loss and causes the brain to focus more on negative aspects throughout the day. This study supports recent findings and offers evidence that the sleep deprived young adults experience impaired visual short-term memory and have an altered perception of the world.
KEYWORDS: Sleep; sleep deprivation; perception; memory consolidation; short-term memory.

■ Introduction

Many people find it hard to remember a name they just heard, a question someone asked, or even where their keys are. Short-term memory loss is often associated with brain tissue damage caused by diseases and certain conditions.[1] Yet, do insufficient sleep and chronic sleep deprivation also have serious effects on the short-term memory capacity of the human brain? Do people who didn't get adequate sleep the other day tend to remember more negative images than positive ones?

Previous research conducted by using repetitive transcranial magnetic stimulation (rTMS) shows that two mechanisms attributable to sleep deprivation have a great impact on short-term memory: perceptual processing and basic attention.[2] Basic attention provides two-channelled receiving functionality by specifically filtering the information for new processes, ultimately leading to perception. The process of choosing, arranging, and evaluating information is known as perception. Perceiving chosen inputs that enter through the sensory filtration, are arranged in established frameworks, and are derived from existing knowledge is what this process entails.[3] Another study concluded that sleep-deprived animals have longer reaction times, and the formation of short-term memory is delayed.[4] Yet, sleep deprivation does not only affect the short-term memory of an individual, but it also causes irritability and lowers the mood.[5] The amygdala, the brain's central processing unit for scary and angry inputs, and the prefrontal cortex, a brain unit important for decision making, executive function, attention, impulsive behaviour, memory, cognitive control, and social behaviour regulation, normally have a balanced relationship. But in a sleepless brain, the amygdala is found to dominate over the prefrontal cortex, which causes a sleep-deprived individual to experience more negative emotions.[6] For instance, in his book "Why We Sleep" Matthew Walker explains that the participants in a study, who normally had the capacity to reset REM sleep at night, but were left without sleep this time, lapsed into a fear-containing bias in a setting where they regarded soft and friendly-looking faces as threatening. The surroundings had become an unjustifiably more terrifying and unpleasant place for them as a result of the brain's lack of REM sleep, a phase of sleep where the eyes move rapidly and randomly. In the perspective of sleep-deprived brains, the actual and perceived reality were not the same anymore.[7]

In accordance with research conducted by many neuroscientists, most of the literature so far demonstrates that sleep deprivation impairs short-term memory consolidation and causes the brain to perceive the world in a negative way.[8] Still, there is hardly any research conducted that precisely examines the sleep deprivation's effects on young adults' visual short-term memory. The purpose of this paper is to offer findings that might give some insight into it, as visual short-term memory is a vital cognitive function for young adults.

The current study aims to help acknowledge the serious influence of inadequate sleep on the human brain's ability to form short-term memories. Specifically, young participants who slept less than others were expected to need a longer time to finish a memory game and likely remember images that evoke negative emotions. This was assessed with the help of a survey.

Firstly, the method of the study and general information about the topic are introduced and are followed by the examination of the results gathered from the memory game and a short-term memory test. The research paper then continues with addressing strengths and weaknesses, meaning of the study, unanswered questions, and future study as well as previous studies and ends with the presentation of the study as a whole.

DOI: 10.36838/v4i4.7

Methodology

Participants:

The study group consisted of young adults. They had to: (i) be aged between 16-18, (ii) live in Turkey or the United Kingdom, (iii) have no sleep disorders, (iv) have no memory disorders, (v) have no vision problems. 106 persons participated in this study. 41.5% of the participants were eighteen, 39.6% seventeen, and 18.9% sixteen years old. 52.8% of those surveyed were female and 47.2% male. All participant young adults were volunteers.

Materials:

A Google Form was created and sent to young adults. The form was divided into three parts: personal information, a memory game, and short-term memory test. The memory game was created using a website called "interacty" and was based on matching ten paired images. The number of minutes each participant spent playing the game was measured by a chronometer. For the short-term memory test, a YouTube video was created. The video showed 20 images for 20 seconds and then required the participants to answer a question. Scores were gathered and then evaluated using various charts.

Procedure:

At the beginning of the survey, the participants were given information about the aim as well as the duration of the survey. Each form could be filled only one time and it was compulsory to fill each question. The survey was sent to young adults in Turkey and the United Kingdom via WhatsApp by me and they needed to fill out the survey at 12 p.m. precisely. They needed to fulfill the participation criteria in order to take the survey. The responses were anonymous so the subjects knew their results on the memory game and the short-term memory test couldn't be traced back to them. After the survey was completed, all the submissions were assessed, and then scatter plots and column charts were created. With the help of suitable charts, the correlation between sleep deprivation and impaired short-term memory was determined. The evaluation process comprised comparing the sleep-deprived participants' memory game durations with others and the type of images (negative or positive emotions triggering) they were able to recall in the short-term memory test. In addition, whether their short-term memory capacity was affected to a certain extent by each decreasing number of hours slept was taken into consideration.

Results and Discussion

Principal Findings:

The purpose of this study was to test whether sleep deprivation had a negative impact on short-term memory consolidation and sleepless young adults were more likely to remember negative images better. I predicted that the longer young adults slept the previous day, the less time they would spend on the memory game (Figure 1 and 2), and the more positive emotions triggering words they would recall on the short-term memory test.[9] The average hours the young adults slept the other day was 6.2 hours. 46.22% of them slept less than 6 hours and 6.6 % slept less than 4 hours the other day. The hypothesis was fully supported for both the memory game and the short-term memory test. The participants who slept less than six hours the other day spent more time on the memory game and remembered more negative emotions triggering images.

Memory Game:

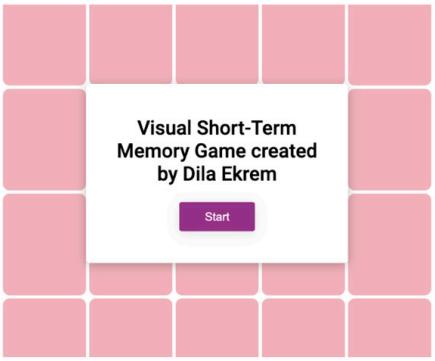


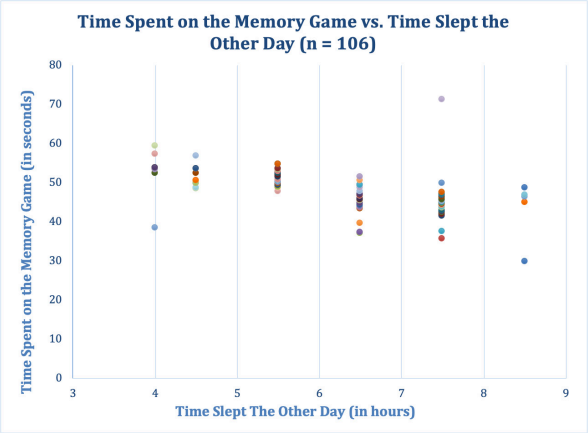
Figure 1: The memory game.

Figure 2: The memory game.

On average, the participants spent around 48.13 seconds on the memory game. The participants who stated having slept more than six hours the other day (n= 57) spent an average of 45.21 seconds on the game, while the participants who stated having slept less than six hours the other day (n=49) spent an average of 51.52 seconds. The participants who slept approximately 8-9 hours the other day (n= 5) spent 43.34 on the memory game on average, indicating that the more the participants slept the less time they spent on the memory game. (Table 1)

Short-Term Memory Test:

Table 1: Results.



DOI: 10.36838/v4i4.7

The participants who slept less than six hours the other day were 74.3% more likely to recall negative emotions evoking images such as the thunder, angry fire, fighting kids, and depressive girl. Around 92.3% of them were able to recall the fighting kids and 88.6% of them recalled the thunder. However, the participants who slept more than six hours the other day were able to recall the fighting kids only 23.4% of the time and the thunder 42.7%. These findings prove that when young adults are sleep-deprived, their amygdala is in control, which explains why they remember negative images better.

Strengths and Weaknesses of the Study:

Strengths and Weaknesses of the Study:

As the study examines the correlation between sleep deprivation and short-term memory loss, the results gathered from the survey strongly support the study's hypothesis. At the same time, as there were 106 participants in total, the results can be generalized to young adults aged between 16-18, who are currently enrolled at a high school and don't have any sleep or memory disorders. In terms of cost-effectiveness, the study was conducted without any costs. The self-compiled survey offered a free method to obtain a huge quantity of data in a short period of time and gather information on a variety of aspects such as personal information, memory/sleep disorders, and sleep habits details. However, if the study was to examine sleep deprivation's effects on short-term memory using neuroimaging techniques, it would be highly costly.

The participants of the survey were requested to fill out the survey at 12 p.m. in order to avoid false results due to the time the survey was taken. Still, the timing might have influenced the results. People can be categorized into three groups when it comes to sleep cycles: morning birds, night owls, and those who identify themselves as both. As 40% of the world population consist of morning birds and 30% of night owls, it can be concluded that 30% of the participants ($n \approx 31$) of the survey were disadvantaged in the sense of executive cognition compared with others.

Previous Studies:

Lisa Y. M. Chuah and Michael W. L. Chee were one of those who studied sleep deprivation's negative effects on short-term memory. Their 2008 research study demonstrates that the reason why short-term memory gets impaired as a consequence of sleep deprivation is that visual attention and visual processing are influenced by it.[10] Their assumptions seem to be plausible when approached with the results of this study.

The findings of this study are also consistent with those of Li, Yu, and Guo's, where they support that sleep deprivation induces a severe deficiency in hippocampus' plasticity in the course of memory formation leading to impaired short-term memory consolidation.[11]

In their cutting-edge paper of 2013, Wee, Asplund, and Chee discovered an impact that is aggravated by sleep deprivation: when short-term memory models are retained over a longer delay, they are more prone to failure.[12] Their results indicate that, if the images on the short-term memory

Meaning of the Study:

Taken together, this study's findings suggest that sleep plays a vital role in the formation of short-term memories among young adults and provides evidence of how sleeplessness can influence the brain's significant functions. The next decade is likely to see many other research studies on sleep deprivation's effects and this study adds substantially to our understanding of the correlation between sleep deprivation and short-term memory capacity.

Unanswered Questions and Future Research:

While the majority of the sleep-deprived participants yielded anticipated results, there were a couple of outliers. For instance, there was in particular a 17-year-old girl who reported having slept less than four hours the other day but spent only 35.6 seconds on the memory game, doing better than most of the participants. This could perhaps indicate that sleep deprivation doesn't necessarily affect young adults' short-term memories in the same way. As in many other aspects of sleep deprivation, whether the short-term memory gets impaired or not might depend on the person. A young adult who had a healthy and quality sleep the other day might show worse short-term memory performance than someone who didn't get sufficient sleep.

Research on sleep's effect on short-term memory and perception of the world could go in various ways. Firstly, studies could explore whether the deprivation of two sleep stages (REM and NREM) has unlike consequences in terms of short-term memory formation. Secondly, the effects of longer sleep-deprivation terms, especially pulling all-nighters, on short-term could be examined. Is there a tipping point, when the brain stops forming short-term memories? Lastly, sleep researchers and neuroscientists could design devices to improve short-term memory formation even when sleep-deprived.

■ Conclusion

I predicted that the participants who slept less than six hours the other day would spend more time on the memory game and would remember more negative emotions triggering images on the short-term memory test. The findings of this study demonstrate my hypothesis through the survey's results, and I found a significant negative correlation between the amount of time slept and the time spent on the memory game. The results of this study have further strengthened my confidence in sleep deprivation's devastating effects on visual short-term memory and perception.

There are numerous examples of sleep deprivation's negative effects on young adults' brains. Taking the results of this study into consideration, the importance of sufficient sleep should be taken more seriously, as the education young adults aged 16-18 get, plays a big role in the formation of their futures, and without adequate sleep, the education they obtain cannot even be useful short-term, let alone long-term. This study, as well as future studies, will add to our understanding of the adverse effects— and potential chronic disorders that can occur — of not getting enough sleep. I hope that this paper successfully highlighted the importance of getting sufficient sleep and will encourage young adults to establish healthy sleep habits to

promote their visual short-term memory. Finally, with the results of this study, it can be concluded that the reason why most people cannot remember where they put their keys is probably that they didn't get enough sleep at night.

Acknowledgements

The completion of this research paper would not have been possible without the support and nurturing of my dear friend Kuzey Özpak, who sent the survey to his friends in the UK and always gave me insightful suggestions during the process. .

References

1. Khan, Z. U.; Martín-Montañez, E.; Navarro-Lobato, I.; Muly, E. C. Chapter One - Memory Deficits in Aging and Neurological Diseases https://www.sciencedirect.com/science/article/abs/pii/B9780124201705000015.
2. Tucker, A. M. Two Independent Sources of Short Term Memory Problems during Sleep Deprivation. *SLEEP* **2013**. https://doi.org/10.5665/sleep.2696.
3. Lumen Learning. What is Perception? | Introduction to Psychology https://courses.lumenlearning.com/msstate-waymaker-psychology/chapter/reading-what-is-perception/.
4. Krishnan, H. C.; Gandour, C. E.; Ramos, J. L.; Wrinkle, M. C.; Sanchez-Pacheco, J. J.; Lyons, L. C. Acute Sleep Deprivation Blocks Short- and Long-Term Operant Memory InAplysia. *Sleep* **2016**, 39 (12), 2161–2171. https://doi.org/10.5665/sleep.6320.
5. Roth, T., Kramer, M., & Lutz, T. (1976). The effects of sleep deprivation on mood. *Psychiatric Journal of the University of Ottawa*, *1*(3), 136–139.
6. Salzman, C. D.; Fusi, S. Emotion, Cognition, and Mental State Representation in Amygdala and Prefrontal Cortex. *Annual Review of Neuroscience* **2010**, 33 (1), 173–202. https://doi.org/10.1146/annurev.neuro.051508.135256.
7. Walker, M. P. Why We Sleep: Unlocking the Power of Sleep and Dreams; Scribner, An Imprint Of Simon & Schuster, Inc: New York, Ny, 2018.
8. Ranganath, C. Working Memory for Visual Objects: Complementary Roles of Inferior Temporal, Medial Temporal, and Prefrontal Cortex. *Neuroscience* **2006**, *139* (1), 277–289. https://doi.org/10.1016/j.neuroscience.2005.06.092.
9. Bell-McGinty, S. Identification and Differential Vulnerability of a Neural Network in Sleep Deprivation. *Cerebral Cortex* **2004**, *14* (5), 496–502
10. Chuah, L. Y. M.; Chee, M. W. L. Cholinergic Augmentation Modulates Visual Task Performance in Sleep-Deprived Young *Adults. Journal of Neuroscience* **2008**, *28* (44), 11369–11377. https://doi.org/10.1523/jneurosci.4045-08.2008
11. Li, X.; Yu, F.; Guo, A. Sleep Deprivation Specifically Impairs Short-Term Olfactory Memory in Drosophila. *Sleep* **2009**, 32 (11), 1417–1424. https://doi.org/10.1093/sleep/32.11.1417.
12. Wee, N.; Asplund, C. L.; Chee, M. W. L. Sleep Deprivation Accelerates Delay-Related Loss of Visual Short-Term Memories without Affecting Precision. Sleep 2013, 36 (6), 849–856. https://doi.org/10.5665/sleep.2710.

Authors

Dila Ekrem is currently a junior at St. George's Austrian High School in Istanbul, Turkey. Her passion for the complex research process together with her curiosity for the human brain inspired her desire to study neuroscience as a major and she is currently interested in sleep, consciousness, and neurological disorders.

DOI: 10.36838/v4i4.7

IJHSR
ijhighschoolresearch.org
■ RESEARCH ARTICLE

Divisibility Algorithm For Number 12

Efe Çete, Fatma Aykaç, Funda Y. Topal
American Collegiate Institute, Göztepe, İnönü Cd. No:476, 35290 Konak/İzmir

ABSTRACT: Divisibility, which can mean dividing an integer by another integer without a remainder, also includes finding the remainder in division. In teaching mathematics; odd, even, prime, etc. states of numbers are important for integers and related subjects, it plays a role in determining many of a number's properties. The divisibility rules of certain numbers are known to students. However, the rules of some numbers are determined by the factors of that number. For example, in the divisibility rule by 12 states, "If the number is divisible by both 3 and 4, it is also divisible by 12." Our aim in this paper is to express the divisibility rule by 12 differently and independently from its multipliers (3 and 4) and to create a completely new algorithm suitable in order to achieve our goal. For our purpose, firstly, it was desired to create a systematic order with numbers that are a whole multiple of 12. Then, a new divisibility algorithm was developed by working with division including remainders. It has been proven that the algorithm works flawlessly in studies using multiple different sized numbers and their resulting mathematical explanation and proofs.

KEYWORDS: Divisibility, digits, remainders, twelve, division with & without remainders.

■ Introduction

"Divisibility" refers to dividing an integer by an integer without a remainder.[1] Divisibility is a subject that contains the information necessary to estimate the remainder in division with remainders, as equally as it is about division without remainders. In the teaching of this subject, the rules are given to the students in the beginning of their studies. Teaching continues by being reinforced with examples. However, in order to improve students' mathematical skills and provide more permanent learning, students can be given numbers that are divisible by an integer without a remainder, and they can be asked to examine whether there is a certain rule between them.[1,2] Thus, the logic of divisibility memorized as a set of rules can be grasped. Yet this would not stay as permanent knowledge without clear explanations rather than straightforward rules, which our paper won't be doing.

When studies on divisibility are examined, there are not many studies on the direct divisibility rules. Based on foreign sources, studies on divisibility rules with prime numbers lead the way.[3,4] The reason for this is that divisibility by a non-prime number is explained by divisibility to all prime factors of the input.[5]

Even though the majority of divisibility rules are prime factor-based, there are some well-known direct rules as well. Certain numbers (such as 2, 3, 4, 5, 7, 8, 9, 10, 11, and 13) have their own divisibility rules that are used by everyone.

For example, A number:

- To be divisible by 2 without a remainder, it must be even.
- To be divisible by 3 without a remainder, the sum of the digits of the number must be a multiple of 3.
- To be divisible by 4 without a remainder, the number in the last two digits of the number must be "00" or a number divisible by 4 without a remainder.

Of course, all these divisibility rules have a proof, a premise. However, even though students know these rules by heart, most of them have no idea why and how these apply.[1] Different activities can be developed to increase students' mathematical reasoning skills of ministry topics and to enable them to produce solution algorithms. In this research, we have examined the number 12, whose rule is accepted such as: "Numbers that are divisible by both 3 and 4 are divisible by 12," and is among the numbers that don't have their own divisibility rules.

Purpose:

During the preparation of the project, our first goal was to answer the question: "Can we find a divisibility rule for any number **other than** the numbers whose rule is known and used by everyone?". In line with the answers we gave to the questions asked, we aim to fill a gap in this field and to the number "12". Just as it was recommended in its teaching, it was started by considering the numbers that are multiples of 12 and their common features.

The study was planned and implemented in order to determine the numbers divisible by 12 and to create an algorithm for finding the remainder and if it is divisible or not.

■ Method

In this project, an algorithm was created to calculate the divisibility of a number by 12 and what the remainder of dividing the number by 12 is.

This divisibility algorithm, which was created on the basis of existing divisibility rules, includes situations such as four operations, steps, and number values. While creating the algorithm, in the first place, a study was carried out on numbers that are a whole multiple of 12 accommodating the "1k-2k" structure of 12, and with the data obtained in "division by 12 without a remainder", it was aimed to reach the conclusion of "division by 12 with a remainder". Continuing on, the answer to the question: "what will be the remainder as a result of dividing a number by 12" was attempted to be achieved. An

DOI: 10.36838/v4i4.8

algorithm has been developed by taking into account the place values of the numbers whose divisibility is investigated through numerous trials. While applying the algorithm, two similar ways are followed depending on whether the number is odd or even.

If the number is "even":

a. The number in the one's digit of the number is separated from the numbers in the other digits.

E.g; A three-digit ABC number is separated from the C number on ones-digit from the three-digit ABC number to obtain the two-digit number AB and the number C. (AB and C)

b. Then, half of the separated number in the ones digit (C/2) is subtracted from the remaining digits (AB two-digit number) (AB - C/2).

c. The number obtained after these operations is divided by 12.

d. If there is a remainder, it is "multiplied by 2 and subtracted from the smallest multiple of 12 bigger than the multiplication result".

Thus, the remainder of the division of the first number by 12 is calculated. If the remainder is "0", the number is divisible by 12 (without a remainder).

If the number is "odd":

a. The number in the one's digit of the number is separated from the numbers in the other digits.

E.g; A three-digit ABC number is separated from the C number on ones-digit from the three-digit ABC number to obtain the two-digit number AB and the number C. (AB and C)

b. Then, "half of 1 more than 1 [(C+1)/2]" of the number in the ones digit is subtracted from the remaining digits (AB two-digit number) [AB-(C+1)/2].

c. The number obtained after these operations is divided by 12.

d. If there is a remainder, it is "multiplied by 2, subtracted from the smallest multiple of 12 bigger than the multiplication result, and subtracted by 1", respectively.

Thus, the remainder of the division of the first number by 12 is calculated.

These steps can be repeated one after the other according to the number of digits of the number we want to calculate its divisibility by 12. It is important to consider the "odd"ness and "even"ness of the remaining numbers after each repetition. A much clearer explanation of the prediscussed topics with examples can be found in sections: 3.1, 3.2, 3.3, 5 and appendices.

Findings:

First of all, we can see that the algorithm works without errors by experimenting with smaller numbers of 3-4 digits. Below are examples where the algorithm is applied only once. Then, while working with larger numbers with multiple steps, the steps of the application are explained according to the changing even and odd situations.

Application of the algorithm on an even number once:

First of all, we can see that the algorithm works without errors byexperimenting with smaller numbers of 3-4 digits. Below are examples where the algorithm is applied only once. Then, while working with larger numbers with multiple steps, the steps of the application are explained according to the changing even and odd situations.

Application of the algorithm on an even number once:

I. The number in the ones digit of the number is separated from the other digits.

→ For the number 3146, the number is divided into two as 314 and 6

II. Half of the number left is subtracted from the remaining number

$314 - \frac{6}{2}$

III. The result obtained as a result of subtraction is divided by 12. (314–3=311)

$$\begin{array}{r|l} 311 & 12 \\ -24 & 25 \\ \hline 71 & \\ -60 & \\ \hline 11 & \end{array}$$

IV. The remainder obtained as a result of division is multiplied by 2.

11 x 2 = 22

V. The result of the multiplication is subtracted from the smallest positive integer of 12 bigger than the result of the multiplication itself.

24 - 22 = **2**

Result of the Operation: The remainder of the division of 3146 by 12 is **2**.

P.S: IV. If the number found as a result of the operation is a multiple of 12, we can say that the number is divisible by 12 without doing V. Operation.

Application of the algorithm on an odd number once:

I. The number in the ones digit of the number is separated from the other digits. For the number 1035, the number is divided into two as 103 and 5.

II. Half of 1 more than the number left is subtracted from the separated digit.

$$103 - \frac{(5+1)}{2}$$

III. The result obtained as a result of the subtraction is divided by 12. (103–3=100)

$$\begin{array}{r|l} 100 & 12 \\ -96 & 8 \\ \hline 4 & \end{array}$$

IV. The remainder obtained as a result of the division is multiplied by 2. 4x2=8

V. The result of the multiplication is subtracted from the smallest positive integer of 12 bigger than the result of the multiplication itself.

12-8=4

VI. In the second operation, "+1" is subtracted from the number obtained as a result of the previous step.

4 - 1=3

VII. Result of Operation: The remainder of the division of 1035 by **12 is 3**.

DOI: 10.36838/v4i4.8

Application of the algorithm on any number multiple times:

I. The number in the ones digit of the number is separated from the other digits. For the number 689814, the number is divided into two as 68981 and 4.

II.Half of the number left is subtracted from the remaining number. (If the reserved number is odd; half of 1 more than the number is subtracted.)

Here the newly formed number is odd. In the next step, half of the number will be subtracted. In this way, single or double cases will be taken into account.

$$68981 - \frac{4}{2} = 68979$$

Operations 1 and 2 are repeated sequentially until a 2-digit number is obtained. (Let's call this operation a step reduction operation.) (We put a "-" sign in order to remind the steps with odd numbers in order not to make mistakes.)

Step 1 — 689814 → 68981 4

$68981 - \frac{4}{2} = 68979$

Step 2 (-) — 68979 → 6897 9

$6897 - \frac{(9+1)}{2} = 6892$

Step 3 — 6892 → 689 2

$689 - \frac{2}{2} = 688$

Step 4 — 688 → 68 8

$68 - \frac{8}{2} = 64$

As a result of repeated operations, the remainder is found from the division of the 2-digit number by 12.

$$\begin{array}{r|l} 64 & 12 \\ \underline{60} & \underline{5} \\ 4 \rightarrow c & \end{array}$$

To make the operations more understandable, we designated the remainder we found as the number c.

IV. Multiply number c by 2.

12 – 4 = 8

V. Subtract the result of the IV.th operation from the smallest multiple of 12 bigger than 2*c.

12 - 8 = 4

The 4th and 5th operations are repeated in reverse (step 4-3-2-1) as the number of steps in the step reduction process.

For the number in step 4 → 4 x 2 = 8

12 - 8 = 4

For the number in step 3 → 4 x 2 = 8

12 - 8 = 4

For the number in step 2(-) → 4 x 2 = 8

12 - 8 (-1) = 3

(Here 1 is subtracted because the number in Step 2 is odd.)

For the number in step 1 → 3 x 2 = 6

12 - 6 = **6**

The number found after these operations gives the remainder of the first number (689814) divided by 12.

Result of Operation: The remainder of the division of 689814 by 12 is **6**.

As you can see, techniques used for different situations are given in the algorithm in the "Method" section. Based on whether the number is odd or even, the divisibility of numbers by 12 and their remaining states are revealed as a result of simple four operations. You can examine Appendix-1 for an example of a number that is exactly divisible by 12, Appendix-2 for the example where the algorithm is applied more than once and all numbers are odd while applying, and Appendix-3 for the example where the algorithm is applied more than once, and all numbers are even while applying.

Mathematical basis of the algorithm:

For clarity, let's take a 3-digit number only. Let this number be ABC. If C is even:

C is separated from the ABC number. If we write this mathematically; ABC= 100A+10B+C= 10.(AB)+C

The number obtained by subtracting C and dividing by 10 will be AB. Then half of C is subtracted from this number AB.

$$\frac{ABC - C}{10} - \frac{C}{2}$$

If we write the equivalent of this expression:

$$\frac{ABC - C}{10} - \frac{C}{2} = \frac{ABC - C}{10} - \frac{5C}{10} = \frac{ABC - C - 5C}{10} = \frac{ABC - 6C}{10}$$

number is obtained. Let's call this last number x. According to the algorithm, this number x was divided by 12 and the remainder was found. Now let's substitute 12k+m for the number x. The goal is to find m, the remainder. This number m would be multiplied by 2 and subtracted from 12 to find the remainder from the division of ABC by 12 (just like in the 4th and 5th steps of the algorithm). Therefore, the remainder should be 12 - 2m.

$$x = \frac{ABC-6C}{10} = 12k + m$$

Let x be this equation. Since the main purpose is the divisibility of ABC by 12, let's write the equation of ABC.

With the product of insides and outsides method:

$$ABC = \mathbf{120k} + 10m + \mathbf{6C}$$

is found. Let's examine this number now. It is clear that 120k is a multiple of 12 here. Since the initially accepted number C is even, it can be seen that C itself contains a factor of 2 and 6C is a multiple of 12. The only imprecise term here is 10m. To ensure this, let's write 10m=12m-2m and rewrite the equation:

$$ABC = 120k + 6C + 12m \mathbf{-2m}$$

The bolded part in the new equation is the remainder of ABC divided by 12. However, since the remainder cannot

be negative, 12 must be added until it becomes positive. So for m<12, the remainder of ABC dividing by 12 is 12-2m. This shows that the algorithm is working correctly. Thus, the conditions for the divisibility of even numbers by 12 are proved.

If C is odd:

According to the algorithm, half of 1 more of C is taken and the operations are repeated. Let C+1=D. As the number is odd (ABC), the divisibility of the even number ABD by 12 is checked. Thus, if the remainder from the division of the ABD by 12 is **y**, it is obvious that the remainder of the division of the ABC by 12 will be **y-1**.

Conclusion

An algorithm has been created in order to perform a divisibility rule by 12 in a different way other than dividing by known and prime factors. Many examples have been studied to show and control how this algorithm we have created works in odd or even numbers. In order to increase clarity, examples of different situations are included in our study in the upcoming sections. In order to find the mathematical basis of our algorithm, which gives error-free results in all cases, a proof is presented with the help of the number values, variables, and division-divisibility information.

As a result, since the algorithm developed allows us to obtain smaller numbers by decrementing the digits in very extensive numbers, it has introduced a brand-new method and a distinct rule for divisibility by 12.

Appendices:

Appendix-1

I. The number in the ones digit of the number is separated from the other digits.

For the number 288, the number is divided into two as 28 and 8.

II.Half of the number left is subtracted from the remaining number.

$$28 - \frac{8}{2} = 24$$

III. The result obtained as a result of subtraction is divided by 12. (28-4=24)

$$\begin{array}{r|l} 24 & 12 \\ \underline{24} & \underline{2} \\ 00 & \end{array}$$

Results

- If the remainder in the division operation is "0" as a result of the 3 sequential operations; It means the number is divisible by 12.
- Thus, we can say that the number 288 is divisible by 12 (without a remainder).

Appendix-2

The algorithm is applied more than once, and all numbers are even while applying:

I. The number in the ones digit of the number is separated from the other digits.

For the number 378784, the number is divided into two as 37878 and 4.

II. Half of the number left is subtracted from the remaining number.

Operations 1 and 2 are repeated sequentially until a 2-digit number is obtained.

i. Step: $378784 \rightarrow 37878 \quad 4$

$$37878 - \frac{4}{2} = 37876$$

ii. Step: $37876 \rightarrow 3787 \quad 6$

$$3787 - \frac{6}{2} = 3784$$

iii. Step: $3784 \rightarrow 378 \quad 4$

$$378 - \frac{4}{2} = 376$$

iv. Step: $376 \rightarrow 37 \quad 6$

$$37 - \frac{6}{2} = 34$$

III. As a result of repeated operations, the remainder is found from the division by 12.

$$\begin{array}{r|l} 34 & 12 \\ \underline{24} & \underline{2} \\ 10 \rightarrow a & \end{array}$$

To make the operations more understandable, we called the remainder we found as the number "a".

IV. Multiply a with 2.

10x2=20

V. Subtract the result of the IVth step from the smallest multiple of 12 bigger than a*2.

24-20 = 4

The 4th and 5th operations are repeated as many times as the number of steps in the IInd step.

iv. → 10x2=20
24-20= 4

iii. → 4x2 = 8
12-8 = 4

ii. → 4x2 = 8
12-8 = 4

i. → 4x2 = 8
12-8 = **4**

- The number found after these operations gives the remainder of the first number discussed by dividing it by 12.
- The remainder of the division of 378784 by 12 is **4**.

Appendix-3

DOI: 10.36838/v4i4.8

The algorithm is applied more than once, and all numbers are odd while applying:

I. The number in the ones digit of the number is separated from the other digits.

For the number 267673, the number is divided into two as 26767 and 3.

II. Half of 1 more than the number left is subtracted from the remaining number.

Operations 1 and 2 are repeated sequentially until a 2-digit number is obtained.

I. Step $267673 \rightarrow 26767 \quad 3$

$$26767 - \frac{(3+1)}{2} = 26765$$

ii. Step $26765 \rightarrow 2676 \quad 5$

$$2676 - \frac{(5+1)}{2} = 2673$$

iii. Step $2673 \rightarrow 267 \quad 3$

$$267 - \frac{(3+1)}{2} = 265$$

iv. Step $265 \rightarrow 26 \quad 5$

$$26 - \frac{(5+1)}{2} = 23$$

III. As a result of repeated operations, the remainder is found from the division by 12.

$$\begin{array}{r|l} 23 & 12 \\ \hline -\,12 & 1 \\ \hline 11 \rightarrow b & \end{array}$$

To make the operations more understandable, we called the remainder we found as the **number b.**

IV. Multiply b with 2.

11x2 = 22

V. Subtract the result of the IVth step from the smallest multiple of 12 bigger than b*2 and subtract by 1.

24-22 = 2,

2 - 1 = 1

The 4th and 5th operations are repeated as many times as the number of steps in the IIrd step.

iv. → 11x2 = 22
24-22 = 2
2-1 = 1

iii. → 1x2 = 2
12-2 = 10
10-1 = 9

ii. → 9x2 = 18
24-18 = 6
6-1 = 5

i. → 5x2 = 10
12-10 = 2
2-1 = **1**

- The number found after these operations gives the remainder of the first number discussed divided by 12.
- The number found after these operations gives the remainder of the first number discussed divided by 12.
- The remainder of the division of 267673 by

Acknowledgements

Mehmet Gökhan Akbaş, Antalya Bahçeşehir Collage Parkorman Campus, Math Educator, gökhan.akbas@bahcesehir.k12.tr

References

1. Altun, M. (2014). *Eğitim fakülteleri ve matematik öğretmenleri için ortaöğretimde matematik öğretimi*. Bursa: Aktüel Alfa Akademi.
2. Baki, A. (2008). *Kuramdan uygulamaya matematik eğitimi*. Ankara: Harf Eğitim Yayıncılıg
3. Smith, F. (1971). Divisibility rules for the first fifteen primes. *The Arithmetic Teacher*, *18*(2), 85-87.
4. Bezuszka, S. J. (1985). A test for divisibility by primes. *The Arith metic Teacher, 33*(2), 36- 38, http://www.jstor.org/.
5. Peretti, A. (2005). Some notes on divisibility rules. *Department of Economics. The University of Verona*, http://dse.univr.it/home/workingpapers/wp2015n19.pdf.

IJHSR
ijhighschoolresearch.org
■ RESEARCH ARTICLE

"I Want It That Way," But Is There a Better Way? The Significance of Numerical Differences in Lineup Accuracy

Faith D. Lee
Strongsville High School, 20025 Lunn Rd, Strongsville, Ohio, 44149, USA; Faith0100@outlook.com

ABSTRACT: False incarceration is the process by which an individual is wrongfully convicted for a crime--often due to bearing a close resemblance to descriptions made in a witness testimony or sketch depiction. The current system of criminal lineups consists of seeing groups of four to five suspects at one time for witnesses to distinguish the true criminal. Differing sizes of lineups has yet to be investigated in relation to accuracy of convictions despite the growing unreliability of this methodology. To test the numerical significance of criminal lineups, survey participants were split into three experimental groups that then had to find the criminal from a simulation out of eight possible suspects. The first experimental group was presented with one-person lineups, the second was shown two-person lineups, and the third experimental group was shown four-person lineups. The rate of accurate conviction was highest in the smallest criminal lineups due to the increased time elapsed during the deliberation period. Lineup size and accuracy rate proved to be inversely related; as each experimental group doubled in size, the accuracy nearly halved in response (from 44.4% to 23.7% to 10%).

KEYWORDS: Behavioral and Social Sciences; Cognitive Psychology; Sociology and Social Psychology; Memory Recall.

■ Introduction

The Innocence Project, a renowned nonprofit that works to overturn false incarceration, stated that eyewitness misidentification [is] the single greatest cause of wrongful convictions nationwide – playing a role in more than 75% of convictions overturned through DNA testing.[1] Acknowledging the discrepancy between accurate identification of criminals and false accusation of innocents allows one to put into question how the criminal justice system can improve. Eyewitness identification refers specifically to the process of criminal lineups. The current system presents groups of four to five suspects at a time in a live setting or as many as eight to ten suspects in photo form. A possible rationale for the implementation of more sizable lineups is efficiency; in theory, this system allows witnesses to evaluate many suspects at once and the ability to use comparison advantageously to correctly identify the criminal.

Although timeliness is especially important to consider for criminal offenses, it needs to be considered that larger lineups lead to a more collective evaluation of suspects instead of placing focus on finding the single criminal. Madore and Wagner state that psychological science and neuroscience indicate that our minds are taxed by multitasking… we must engage in task switching, placing increased demands on neurocognitive systems… [and] some performance deficit typically occurs.[2] The deficit reveals itself in low accuracy rates of criminal lineups and high false incarceration rates. This occurs because, in a large group scenario, a witness no longer spends their time observing everyone separately. Alternatively, they subconsciously weigh each suspects' traits in reference to others--blurring the focus needed to convict a single criminal.

The belief that larger lineups are objectively better for the witness also conflicts with an aspect of memorization research known as the chunking method. The chunking method is the process of breaking down large amounts of information into smaller, logical units that are easy to understand. Chunking is a valuable tool for memorization.[3] This research alludes to the fact that observing smaller fractions of suspects at once may enhance memorability and ease for witnesses as they attempt to recall the incident. One experiment defines exactly why that simple change alleviates stress, saying that a chunk reduces the load on working memory (WM) via retrieval of a compact chunk representation from long-term memory that replaces the representations of individual elements of the chunk.[4] There is something to be said about considering a witness's perspective; after witnessing a possibly traumatic event, they may feel numerically overwhelmed and unfamiliar with viewing the suspect in a group setting. Seeing individuals one-on-one, emulating the interaction made upon the crime scene, may spark familiarity and therefore allow for a more ideal environment for memory recall.

Conversely, smaller lineups may pose a dilemma, and consequence, like that of the Monty Hall Problem. Contestants on the show "Let's Make a Deal" were informed that behind one of [three] doors there is a car while behind the other two there are goats. [5] Once a player selected a door, the host, Monty Hall, would open one of the remaining two doors to reveal a goat, and he would allow the contestant to switch. Despite the fact that switching doors leads to a ⅔ probability of choosing the car–in contrast to the initial door's probability of ⅓–the vast majority of people show a strong tendency to stick with their initial choice.[6] This phenomenon can be justified through "cognitive illusion", often used to demonstrate people's resistant deficiency in dealing with uncertainty,[7] and the larger amount of regret participants anticipate to experience after a loss due to switching rather than to a loss due to staying.[6] The Monty Hall Problem shows a resistance to opening more doors, or in this case viewing

DOI: 10.36838/v4i4.9

more lineups, after an initial choice is made, which could negatively affect the accuracy of small criminal lineups.

Methods

Survey Layout and Consent:

Three experimental groups of high school participants were asked to take an anonymous interactive survey. This survey was intended to simulate the witnessing of a crime and the criminal lineup process. Participants were directed to watch a video in order to comply but were not informed why they needed to do so or how important it would be to the experiment--similarly to how suspects are not warned prior to viewing a crime that they need to pay attention. After students viewed a staged video of a theft, they answered demographic questions about themselves in order to create a buffer between viewing the video and the need for memory recall. Students were then informed that they needed to think back to the criminal from the video and accurately identify the prime suspect out of 8 individuals. Reasoning questions about each suspect appeared next to facilitate meaningful judgment in evaluating whether each suspect was innocent. Following this, students were asked whether each suspect was guilty. The survey would end when participants convicted a suspect, but if they continued through the entire survey and were unable to find them, they were redirected to a special page* (see experimental group section).

Participants were informed prior to experimentation that the simulation contained mention of possibly triggering topics i.e., conviction, crime, criminal justice, etc. Though the simulation was clearly staged, appropriate to view, and approved by a local Institutional Review Board, the decision of consent remained for everyone. A theft was chosen in order to avoid emotional harm to even those unaffected by crime--hence why there existed no mention or imagery of gore, violence, threatening behavior, or even direct contact. Theft is not a crime that directly disturbs bystanders, but it remains an easily identifiable illegal action to be observed.

Experimental Groups:

In order to test the numerical significance of differing lineup size regarding accuracy, three different experimental groups were created. At the very start of the memory recall process, participants were informed that there existed eight total suspects to choose from. Though the same suspect pictures were shown in the same order of evaluation, the lineup size differed within the three groups. The first experimental group was presented with a one-person lineup, the second was shown pairs, and the third experimental group had four person lineups. Participants were randomly assigned to a group by three administering teachers.

The creation of live lineups was not possible in person due to the limitations of COVID-19, but survey pages made it possible to control what was shown at once. For example, experimental group two was only able to see two suspect pictures at the top of each page; if they asked for another page of the survey, it would be likened to asking for another lineup to be presented in real life. Google Forms allows owners to enable restrictive movement through pages so that participants would be unable to continue onto the next page unless all questions from the previous page were filled out. With this precaution in place, there existed no chance that students were able to view lineups ahead of time or view remaining suspects that they were not supposed to see to ensure validity of results and strict lineup protocol.

After a conviction was made, participants were brought to an ending screen. This page included an announcement that the experiment was over, a reminder that the suspects or simulation were not real, and an optional question asking them to briefly explain why they chose the suspect.

There existed a population of participants who continued through the entire experiment, completed all their lineups, and stated that they were unable to find the criminal. These individuals were redirected to special page* where participants were given a single picture with all eight suspects shown at once. They were then asked, in hindsight, if they could find the criminal. This was an additional test of an extremely large lineup of eight at a time but differed from other experimental groups because not all individuals were given the opportunity to see this unless they were unable to find the criminal. The deliberation period between choosing the criminal was also extended since the individuals--regardless of what experimental group they began with--were estimated to have objectively spent the most time contemplating their decision. This is a valid prediction to make as other participants who chose a criminal ended their survey early due to not being able to see other options.

Survey Layout and Consent:

An important aspect of the simulation was a video of a crime that the students would be required to view. I decided to use my face as the true criminal in the video due to the limitations COVID-19. It was deemed a dangerous task to invite another individual to film the video with myself present at the time of experimentation, especially due to the necessity of being unmasked: facial recognition relies heavily upon the observation of expression, jaw structure, and specifics of the nose and mouth--of which the mask conceals. Additionally, it is to be noted that certain individuals that were consulted found the possibility of their face being seen as a guilty person or possible criminal outside of the experiment worrisome. Although it was made clear that the simulation was a staged production, the fear remained within participants that their face would be seen in a negative light by survey takers. This was a risk I was willing to take for my research to proceed but remained yet another justification as to why my face was the most safe and ideal choice for the prime suspect.

The only concern that could arise with the usage of my face as an integral part of the survey was the possibility of internal bias. There existed a chance that I would portray myself differently than how I truly look or create versions of myself that were indistinguishable--due to a hyper awareness of my specific features. In order to combat this, I had all media with my face checked for objective accuracy by an Institutional Review Board and created the suspect picture of my face using the artificial intelligence website called Artbreeder* (see Artificial Intelligence Photographs).

only outlier (of having large lineups produce less accurate results) was seen in the special page.

The accuracy rate of this population overall was 60.1%. Participants who were sent from seeing lineups of one person at a time were correct 69.2% of the time in contrast to those sent from lineups of two at a time who were correct exactly 50% of the time. Zero students from the lineups of four elected to continue searching in the special page. The pattern of increasing accuracy increased with smaller lineups, as aforementioned, but it was even higher for small lineups who went to the special page.

Conclusion

The data suggests that the defining factor of this change was the amount of time spent in the deliberation process. For the individuals who went through the tedious and slow process of only seeing one person at a time, there existed extended time spent waiting for each lineup. There was an increased attention to detail as shown in the last question which asked for a brief justification for participants' final verdict. The respondents did not say their main reason was instinct, like those in larger lineups, but rather mentioned specific features they recognized (i.e., shape of eyes, young skin, width of face, facial proportions, etc.). The special page surveyors also devoted a great deal of time because they continued searching and persevering through every single participant. Thoughtful consideration is cultivated in these extended time periods and moments of reflection. The other two large experimental groups most likely spent less time truly observing each individual and more time comparing generalities between the lineups. This draws focus away from the moment's recognition in witnessing a crime and instead leads to more decisions being made from bias.

If it remains true that increasing the amount of time spent in deliberation allows accuracy rates to increase as well, my conclusion implies the possibility of a safer society. The implementation of smaller lineups or an extensive period of seeing extremely large lineups could decrease the amount of false incarceration because witnesses can truly utilize facial recognition at its highest accuracy rate. This is a costless and objective means of police reform that could imply a change in a long unrevised system in America.

Acknowledgements

Sincere thanks to Mrs. Michelle Borelle for being an inspirational correspondent and for her generous time lent in advising the early stages of my methodology. None of this could have been possible without the help of Mrs. Alyssa Hoslar, my teacher and supporter always. Her devotion to fielding my questions and her valuable critiques allowed me to produce the experiment to my utmost potential. I wish her and her future capstone students well in AP Research.

References

1. Innocence Project. One Suspect, One Crime, Two Police Sketches. https://innocenceproject.org/one-suspect-one-crime-two-police-sketches/ (accessed October 10, 2020).
2. Madore, Kevin P.; Wagner, Anthony D. National Center for Bio technology Information. Multicosts of Multitasking. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7075496/ (accessed October 26, 2020).
3. University of St. Augustine For Health Sciences. Science-Backed Memory Tips and Recall Techniques. https://www.usa.edu/blog/science-backed-memory-tips/ (accessed November 11, 2020).
4. Thalmann, Mirko; Souza, Alessandra S.; Oberauer, Klaus. National Library of Medicine. How Does Chunking Help Working Memory? https://pubmed.ncbi.nlm.nih.gov/29698045/ (accessed November 9, 2020).
5. Gill, Richard D. Mathematical Institute, University of Leiden. The Monty Hall Problem. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.297.2144&rep=rep1&type=pdf (accessed February 17, 2022).
6. Saenen, Lore; Heyvaert, Mieke; Van Dooren, Wim; Schaeken, Walter; Patrick Onghena. National Library of Medicine. Why Humans Fail in Solving the Monty Hall Dilemma: A Systematic Review. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6194549/#B1 (accessed February 18, 2022).
7. Krauss, Stefan; Wang, X.T. Journal of Experimental Psychology. The Psychology of the Monty Hall Problem: Discovering Psychological Mechanisms for Solving a Tenacious Brain Teaser. https://www.researchgate.net/publication/10839512_The_Psychology_of_the_Monty_Hall_Problem_Discovering_Psychological_Mechanisms_for_Solving_a_Tenacious_Brain_Teaser (accessed February 17, 2022).
8. Simon, Joel, Artbreeder, version 2021, 2021.

Authors

Faith Lee is a senior at Strongsville High School impassioned with the subjects of behavioral and cellular science. She plays contemporary and classical cello in her free time in addition to writing poetry, volunteering, and contributing as an active member in honors societies at her school. Faith has chosen to continue her education and research journey at The University of Michigan this fall.

DOI: 10.36838/v4i4.9

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Cancer mRNA Vaccines as a Promising Approach for Treating Luminal A Breast Cancer

Alexandra Giuliani
International School of Monaco, 12 Quai Antoine 1er, Monaco, Monaco, 98000, Monaco; alexandra.giuliani@yahoo.com

ABSTRACT: Luminal A breast cancer is the most common subtype of the most common cancer in women. Current treatments using surgery and chemotherapy have greatly improved patient outcomes in recent decades but are still far from perfect. mRNA vaccines, which instruct the body to create specific proteins, hold the best promise going forward as a "cure" for luminal A breast cancer. Immunological therapies are emerging in solving problems that chemotherapy, surgery and targeted therapies cannot. The ultimate goal of immunotherapy would be to help the immune system recognize and destroy tumor cells as well as make antigens. In this paper, mRNA's potential for addressing the heterogenous nature of tumors, envisioning how it might be made to effectively elicit an immune response are explored; in particular, how it can be modified quickly to change which mutated genes it targets. This paper also discusses the limitations of this emerging technology and the difficulties of evaluating its efficacy when it has yet to enter clinical trials. Finally, this paper concludes by offering thoughts on how its development might be accelerated, predictions on effective targets, and likely dates for such a vaccine to debut as a real-world treatment.

KEYWORDS: Biomedical and Health Sciences; Cellular and Molecular Biology; Cell, Organ and Systems; Immunology; Cellular Immunology; Oncology; Breast cancer; mRNA.

Introduction

Breast cancer is the most common type of cancer in women besides skin cancer and the second most common cause of death after lung cancer. Worldwide, there are about 1.7 million cases of breast cancer diagnosed every year, with approximately one new case detected every 18 seconds.[1] It is estimated that 281,550 women in the US will be diagnosed with invasive breast cancer and 49,240 women will be diagnosed with non-invasive breast cancer this year.[2] There are several factors that may raise a woman's risk of developing breast cancer such as increasing age; personal history of breast cancer; family history of breast cancer; nongenetic, Nonmodifiable Risk Factors like race, early menarche /late menopause, etc.; and Modifiable Risk Factors like hormone use, tobacco, alcohol, and nutrition.[3]

Breast cancer begins when the healthy cells in the breast grow out of control and form a mass or sheet of cells called a tumor. Mutations affect the function of tumor-suppressor genes and/or oncogenes. The 'multiple hit model' of cancer formation hypothesizes that a single cell must have several damaged genes and receive a series of mutations that build up over time in order to become cancerous. This model shows why cancer is so hard to treat: these multiple genes mutate and interweave to create a complex and heterogeneous population of tumor cells.[4] In a sense, one tumor can be considered multiple diseases. This is why scientists are trying to harness the immune system to fight cancer—it is the best possible defender against such a varied onslaught. Tumor suppressor genes can slow down cell division, repair damaged DNA and make cells go into apoptosis (programmed cell death). Cancer can be formed when these tumor suppressor genes do not work properly, causing the cells to grow out of control.[5] Proto-oncogenes aid with cell growth; however, when these genes mutate/change or overproduce copies, they can become permanently activated, from (proto-oncogene to oncogene) when they are not supposed to be. This causes the cell to grow out of control, leading to cancer.[6]

In this paper the breast cancer subtype Luminal A, also known as HR+/HER2- will be the focus. It is the most common subtype, making up 68% of all breast cancers.[7] Hormone receptor (HR) positive means that the tumor cells have receptors for the hormones estrogen or progesterone, leading to the development of HR+ tumors. Human epidermal growth factor receptor 2 negative (HER2-) means that the tumor cells do not have abnormal levels of HER2 proteins.[8]

Prognosis for luminal type A is better than other breast cancer subtypes since it has high hormone receptor expression, negative HER2 expression (does not grow fast and is not likely to spread to the lymph nodes quickly) and a low proliferation rate (slowly dividing cells).[6] While these factors give those with this cancer a better prognosis than most, it is notable that this subtype is the most commonly diagnosed. This combination of factors makes it a good candidate for testing experimental treatments to demonstrate the efficacy of new cancer technology. Treatment for breast cancer represents a great strain on the resources of public health, so essentially eliminating it would act as a boon to the system with cascading benefits for everyone.

Current Treatments for Luminal A Breast Cancer and Their Shortcomings:

Treatment decisions for breast cancer rely on immunohistochemistry markers as well as nodal status, tumor grade, and tumor size. Depending on these factors, one or a combination of these three treatment routes will be decided upon:

DOI: 10.36838/v4i4.10

Surgery:

Early-stage breast cancers are often easily treatable with surgery, but undetected and large tumors progressing to later stages quickly become more difficult to treat this way.[9] (This is because of the difficulty inherently involved in determining whether all cancer cells of large tumors have been successfully removed) Surgery is performed to remove as much cancer as possible. A mastectomy is a type of surgery where the entire breast or both breasts (double mastectomy) are removed. In contrast, a lumpectomy/breast-conserving surgery is one in which only the specific part of the breast with cancer is removed. An advantage of this type of surgery is that a woman can keep most of her breast, though she will often need radiation therapy as well. For early-stage breast cancer, breast-conserving surgery with adjuvant radiotherapy works best. Navigating the timeline (immediately, 6 months after, etc.).[9] of and complications around reconstructive surgery adds an additional element of difficulty with this treatment pathway.

Chemotherapy:

Chemotherapy controls the cancer by restraining its spread, making it grow more slowly and killing cancer cells that may have metastasized (When the cancer has spread from one part of the body to another, presenting further complications to treatment that must move beyond the localized). Neoadjuvant chemotherapy (before surgery) might be given to patients to shrink the tumor so that it may be removed with less extensive surgery. Usually used to treat locally advanced cancer, neoadjuvant chemo can lower the risk of recurrent cancer. Adjuvant chemotherapy (after surgery) is given to kill cancer cells that were not removed in surgery or that might have spread and cannot be seen. This also lowers the chance of recurrent cancer. Chemo is often most effective if combinations of drugs are used: anthracyclines such as Adriamycin and Ellence can be utilized alongside taxanes such as Taxol and Taxotere for both adjuvant and neoadjuvant chemo, while more aggressive anthracyclines like Doxorubicin, pegylated liposomal doxorubicin, and Epirubicin are combined with taxanes like Abraxane to treat metastasized breast cancer. Side effects of these drugs include menstrual changes and fertility issues, heart damage, nerve damage (neuropathy), hand-foot syndrome, chemo brain, and fatigue.[10]

When looking specifically at luminal A and how it reacts with chemo, studies fail to show benefit for patients. It still remains to be determined if chemotherapy has clinical significance, especially for patients with positive lymph nodes—meaning that the cancer has spread from the original tumor to the surrounding areas but has not metastasized yet. One review compared 5-year survival-rate differences among patients with Luminal A tumors who received and did not receive chemotherapy: the rate for patients at high clinical risk and low genomic risk without distant metastases who received chemotherapy was 1.5% higher, while the absolute difference among patients at low clinical risk and high genomic risk was only 0.8% higher.[11] As nearly all luminal A-like tumors fall within the low genomic risk category, these results demonstrate little benefit of adjuvant chemotherapy with this cancer subtype. Additionally, studies have shown that high-risk premenopausal patients with other breast cancer subtypes had beneficial results from receiving chemotherapy, while Luminal A patients did not experience such gains. Patients may even be harmed by the toxicities of the chemotherapy.

Targeted Therapies:

The main shortcoming of targeted therapies is that they are still very early in development. They rely on administering drugs specific to the patient's cancer subtype, generally in combination with other treatments, to perform actions like tumor shrinkage.[12]

For decades, endocrine therapy (fulvestrant-FASLODEX, letrozole-FEMARA) has been the cornerstone for management of luminal breast cancer. Despite the substantial benefit derived by patients from endocrine therapy, primary and secondary resistance to endocrine therapy are serious clinical issues.[13]

Pathways involved in the biology of endocrine resistance have been well studied in the last two decades, leading to the development of several classes of targeted agents that have been approved. Today, in the advanced setting, three distinct classes of targeted agents are approved for use: mTOR (everolimus-AFINITOR), CDK 4/6 (palbociclib-IBRANCE, ribociclib-KISQALI, abemaciclib-VERZENIOS) and PI3K inhibitors (alpelisib-PIQRAY).[14]

CDK 4/6 inhibitors are the most important of these, having changed the natural history of this disease in the advanced setting and being currently under study in the early setting.[15] In short, combining endocrine and targeted therapies has changed the landscape in advanced disease; in early disease, it is possible to have a similarly large impact, particularly in patients with higher risk of relapse. Moreover, experimental targeted drugs are in development such as AKT/PTEN inhibitors (Ipatasertib and capivasertib).[16]

Efforts to develop new agents with SERD (Selective Estrogen Receptor Degrader) properties with potent antiestrogenic activity in breast tissue have led to the discovery and characterization of second and third generation SERDs, orally bioavailable, some of which are now undergoing clinical evaluation. These include Elacestrant, SAR439859, GDC-0810 and AZD9496, among others.[17]

Finally, it is noteworthy that the development of targeted-agent combinations for luminal disease faces several challenges. Firstly, outcomes are already exceedingly good for most patients. Additionally, late recurrences—after five years—are a possibility for at least 20 years, which translates into a very long period of follow-up looking for events to occur. Lastly, side effects of treatment considered acceptable in advanced disease, such as such as alopecia, may not be tolerated by patients with early disease.[18]

The Potential of mRNA Vaccines for Treating Luminal A Breast Cancer:

While there are many potential new treatments being explored for cancer, such as CRISPR gene-editing, microbiome treatments, and cell therapy, the one that holds the most prom

DOI: 10.36838/v4i4.10

ise for luminal A breast cancer in particular is mRNA vaccine technology. There are a few types of vaccines being explored as potential cancer treatments: immune cell-based vaccines, peptide-based vaccines, viral vector-vaccines and nucleic acid-based vaccines. (At least one vaccine, inoculating teenagers against the human papillomavirus, is currently working to reduce future rates of cervical, anal, oropharyngeal, etc. cancers that are caused or contributed to by the virus.)[19] mRNA vaccines in particular are classified as nucleic acid-based and hold an especial potential to protect against infectious diseases and rapid malignant cell growth.

mRNA is, in essence, a template of instructions for how to build a protein. This technology has been gaining promise as the coronavirus pandemic has brought a lot of attention to it. The two most prominent COVID-19 vaccines in the US both utilize mRNA, which has skyrocketed interest (and funding) for additional mRNA research.[20] Both BioNTech and Moderna are pursuing mRNA cancer vaccines.[21] These prove to be a little trickier than an mRNA vaccine for a virus since that case deals with clearly foreign viral matter, while cancer vaccines have to train the body to identify those portions of itself that are "foreign" malignant cells.[22]

How will this work? A particular sequence of mRNA will be injected into the muscle and taken up by any cells in the surrounding area. The immune system will then hopefully recognize the proteins coded for by the mRNA as foreign and mount an immune response. The result of the immune response is that if the antigens are immunogenic enough, the T-cells (white blood cells that actively seek out the cancer cells and destroy them) will learn how to recognize these cancer cell antigens as foreign. This chain of events should trigger an immune response, wherein B cells (which produce antibodies that protect us from getting infected if, for instance, a virus enters the body) will subsequently seek out similar tumor cells.[23]

mRNA-based cancer vaccines can target tumor-associated antigens, expressed in cancerous cells, for example, growth-associated factors, or choosing specific antigens that are unique to malignant cells owing to somatic mutation. These neoantigens provide tumor-specific targets for developing personalized cancer vaccines.[24]

In short: upon administration of mRNA vaccines, tumor antigens will be expressed in a heterogenous group of immune cells in order to help antigen-presenting cells activate and elicit an immune stimulation.

To ensure the mRNA reaches its destination, naked mRNA (after *in vitro* transcription, it is not bound to protein and vulnerable to different enzymes that are going to recognize it) can be formulated with a delivery vehicle (vehicle-loaded mRNA).[25] Delivery vehicles help to improve stability, RNA uptake and translatability of mRNA vaccines. Application of exogenous RNA combined with a polymeric carrier was also shown to activate the immune system by generating a local immunostimulatory environment.[26]

mRNA vaccines are a strong choice over other vaccines for treating cancer because they are potent, safe and easily (as well as inexpensively) modifiable. The production of mRNA is simpler and easier than protein production and purification and results in a more stabilized product.[27] The trick is in trying to deliver the mRNA in a way that helps maximize potency while doing that safely. Potential methods for maximizing potency include modifications made to the structure of the mRNA molecule (e.g., self-amplifying mRNAs, codon optimizations, nucleotide modifications, etc.) and different means of formulation (lipid nanoparticles, peptides, polymers, etc.).[27] But again, a strength of working with mRNA is how readily its sequence can be changed to add or swap targeted antigens. The quickly swappable nature of mRNA vaccines is the best way to address the challenge presented by tumors' high degree of heterogeneity.

While mRNA vaccines are providing a powerful new avenue for treating cancer, there do remain challenges to creating an effective mRNA cancer vaccine. The products are required to be the right purity of mRNA, sequenced as well as delivered and administered properly. Given the difficulty of determining what factors are specifically driving an individual's cancer to be out of control, it is hard to pick the right protein to encode instructions for in the mRNA that will elicit the needed immune-system response. This goes alongside the general issue with cancer therapies of managing the differentiation of normal tissue from cancer cells.[27]

That said, this work is not merely theoretical: mRNA vaccines are already being developed in the treatment of other cancers (such as skin, lung, and pancreatic cancers).[28] Several clinical trials with mRNA vaccines against cancer are ongoing, and nonclinical research is active in this field. Noteworthy results were released in November 2020 of an mRNA mixture being injected into mice with skin and lung cancer that successfully triggered an immune response. Cytokines produced by this response were able to shrink 85% of tumors in the mice in the span of 40 days.[29] These cancer vaccines, produced by the same companies behind the mRNA vaccines for COVID-19, will soon be entering clinical trials with humans in an exciting next step.

Possible Components of an mRNA Vaccine for Luminal A Breast Cancer:

In light of the above, a theoretical combination of treatments that will in effect "cure" luminal A breast cancer can be proposed. In particular, a way forward using mRNA vaccines that will effectively address the inherent heterogeneity of tumors that so often impedes traditional cancer treatments can be outlined.

From the outset, picking the right target is the biggest hurdle in making the mRNA vaccine against cancer. The first potential target would involve picking a protein present in the cancer cells but not the healthy ones—i.e., the genes that are most mutated. When looking at luminal A breast cancer, PI3KCA is the most mutated gene (about 45%), followed by MAP3K1, GATA3, TP53, CDH1, and MAP2K4.[30] Genes like PI3KCA contain instructions to make proteins such as the p110 alpha protein (p110-α, subunit of phosphatidylinositol 3-kinase). This gene encodes a lipid kinase involved in vital signaling pathways, fundamental for cellular functions (e.g., growth, death, and proliferation).

However, when PI3KCA is mutated, there is a change in the single amino acid produced in the p110-α, leading to the production of an altered p110-α subunit that in turn makes PI3K abnormally active. This allows PI3K to signal without regulation. The unregulated signaling results in uncontrolled proliferation of cells, causing cancer.[31] Therefore, targeting the altered p110-α in this mRNA vaccine could successfully short-circuit this entire process.

To circumvent cancer resistance observed with chemotherapy and given the heterogeneity of tumors (not every cancer cell they are in will be equally vulnerable to the use of one targeted antigen), it would be better to use two or more mRNAs. As such, this vaccine could ideally deliver not one but several different mRNA sequences encoding for distinct proteins, ideally eliciting a robust immune response towards the full spectrum of what the cancer is expressing. However, a possible drawback here is the potential for creating a situation where there is an autoimmune response.[25] It is not desirable to encourage the immune system to attack more than just the cancer and damage healthy cells, and yet more targets inherently increases the chances for this to happen. To avoid this scenario, it may make sense to start with a single mRNA molecule and test for whether it creates a strong enough response; if it does not (as is likely), a combination of molecules would be used or one after the other could be used to see what works best. A combination of PI3CKA, MAP3K1, and GATA3 could work well, for instance, at which point it would be possible to pare down to fewer targets to determine which target or combination of targets made it so effective.

Direct administration of complex mRNA is now considered to be a fast and feasible approach. However, there are several problems with the mRNA vaccine such as limited *in vivo* delivery, possibly due to enzymatic degradation, and limited intracellular delivery due to the large size of the mRNA molecule compared to other payloads. This could be overcome by screening different delivery materials and formulation methods.

Another possible avenue of interest might be to combine mRNA vaccines with immune checkpoint inhibitors. Most tumors strive to evade the immune system by expressing immunosuppressive proteins on their surface, called checkpoint proteins. Immune checkpoint inhibitors work by blocking that checkpoint. This prevents the "off" signal from being sent, allowing the immune cells to kill cancer cells. For instance, high CD70 expression may inhibit the anti-tumor immune response and is a promoter in tumor progression.[32] CD70 expression (how much protein there is) could allow the tumor cells not to trigger an immune response. By inhibiting immune response, the CD70 could work against the mRNA vaccine. As such, any ways to diminish that possibility (such as combining with different treatment options) should make this vaccine more effective.

Discussion

Only studies can truly determine the efficacy of new treatments, which limits the possibilities of what a literature review in this area can accomplish. And as few mRNA vaccines for cancer treatment are yet in clinical trials, they are not widely reported on in the literature—making research into this area difficult. That said, the success of mRNA vaccines in fighting the spread of the novel coronavirus signals a clear turning point in this technology. The possibility of a groundbreaking treatment here should not be denied.

With the right antigen selection and fine-tuning of mRNA delivery, payload size, etc., it is possible to unlock the full potential of the best ally each human body always had in attacking cancer cells: the human immune system. If the immune system can be properly instructed to treat cancer cells as the "foreign" bodies that they essentially are, their eradication of luminal A breast cancer tissue could end up being far more complete than what is possible with surgery and chemotherapy alone—and in a way that improves patients' physical and psychological outcomes as treatment becomes less invasive.

This 'hacking' of the immune system is a new and exciting field of science that is moving at a rapid clip. With how quickly mRNA vaccines can be developed, deployed, redeveloped, and redeployed, progress should be expected quickly. The promise of mRNA vaccines (and possibly oRNA and miRNA as well) as proposed above for luminal A breast cancer is not a far-off or far-flung dream, but a promise that will likely be realized within the next decade or two.

Conclusion

mRNA vaccines are a highly promising route for treating or even "curing" luminal A breast cancer. They should continue to be explored at an aggressive pace, with increased funding allotted to speed them into clinical trials in the wake of their massive success in combating COVID-19.

It is likely that, in order to effectively address the heterogeneity of tumors, these vaccines will have to target multiple antigens simultaneously. This paper predicts that PI3KCA and MAP3K1 will prove the most worthwhile genes to target, and that doing so with effective payload delivery and minimal autoimmune response triggered will be possible in real cancer therapies by 2035. Particular attention should be paid to how best to combine these mRNA vaccines with existing treatments like surgery and chemotherapy as well as other cutting-edge treatments like targeted therapies.

These advances in treatments should function as a "cure" for luminal A breast cancer by 2040, with the relieved burden on the health care system as well as on women everywhere ushering in a new age of wellness and positive life outcomes. Even more excitingly, given the flexibility with which researchers can rework mRNA in the lab, any advances in mRNA vaccines for cancer should prove a ready boon to cancer researchers of all stripes—and assuredly to those looking into other perennial diseases likewise.

Acknowledgements

Thank you for the guidance of Daniel Wilson, my mentor from Carnegie Mellon University, in the development of this research paper.

References

1. Breast cancer statistics https://www.wcrf.org/dietandcancer/breast-cancer-statistics/.

DOI: 10.36838/v4i4.10

2. Breast Cancer - Statistics https://www.cancer.net/cancer-types/breast-cancer/statistics.
3. CDCBreastCancer. What are the risk factors for breast cancer? https://www.cdc.gov/cancer/breast/basic_info/risk_factors.htm.
4. How common is breast cancer? https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html.
5. Harbeck, N.; Penault-Llorca, F.; Cortes, J.; Gnant, M.; Houssami, N.; Poortmans, P.; Ruddy, K.; Tsang, J.; Cardoso, F. Breast Cancer. Nat. Rev. Dis. Primers 2019, 5 (1), 66.
6. Oncogenes and tumor suppressor genes https://www.cancer.org/cancer/cancer-causes/genetics/genes-and-cancer/oncogenes-tumor-suppressor-genes.html.
7. Female breast cancer subtypes - cancer stat facts https://seer.cancer.gov/statfacts/html/breast-subtypes.html.
8. Molecular subtypes of breast cancer https://www.breastcancer.org/symptoms/types/molecular-subtypes.
9. Breast Cancer Surgery https://www.cancer.org/cancer/breast-cancer/treatment/surgery-for-breast-cancer.html.
10.A Chemotherapy for Breast Cancer https://www.cancer.org/cancer/breast-cancer/treatment/chemotherapy-for-breast-cancer.html. .
11.Gao, J. J.; Swain, S. M. Luminal A Breast Cancer and Molecular Assays: A Review. Oncologist 2018, 23 (5), 556–565.
12.Targeted Therapy https://www.nationalbreastcancer.org/breast-cancer-targeted-therapy. Chemotherapy for Breast Cancer https://www.cancer.org/cancer/breast-cancer/treatment/chemotherapy-for-breast-cancer.html.
13.Presti, D.; Quaquarini, E. The PI3K/AKT/MTOR and CDK4/6 Pathways in Endocrine Resistant HR+/HER2- Metastatic Breast Cancer: Biological Mechanisms and New Treatments. Cancers (Basel) 2019, 11 (9), 1242.
14.Fan, W.; Chang, J.; Fu, P. Endocrine Therapy Resistance in Breast Cancer: Current Status, Possible Mechanisms and Overcoming Strategies. Future Med. Chem. 2015, 7 (12), 1511–1519.
15.Spring, L. M.; Wander, S. A.; Zangardi, M.; Bardia, A. CDK 4/6 Inhibitors in Breast Cancer: Current Controversies and Future Directions. Curr. Oncol. Rep. 2019, 21 (3), 25.
16.Oral Selective Estrogen Receptor Degraders (SERDs) as a Novel Breast Cancer Therapy: Present and Future from a Clinical Perspective. International Journal of Molecular Sciences 2021.
17.Cristina Hernando, Belén Ortega-Morillo, Marta Tapia, Santiago Moragón, María Teresa Martínez, Pilar Eroles, Iris Garrido-Cano, Anna Adam-Artigues, Ana Lluch, Begoña Bermejo, and Juan Miguel Cejalvo.
18.Martorana, F.; Motta, G.; Pavone, G.; Motta, L.; Stella, S.; Vitale, S. R.; Manzella, L.; Vigneri, P. AKT Inhibitors: New Weapons in the Fight against Breast Cancer? Front. Pharmacol. 2021, 12. https://doi.org/10.3389/fphar.2021.662232.
19.Gallagher, J. HPV Vaccine Cutting Cervical Cancer by Nearly 90%. BBC. November 4, 2021. Vinluan, F. Can MRNA Vaccines Do for Cancer What They Did with Covid-19? MedCityNews. 2021.
20.Deutsche Welle (www. dw.com). Will there be an mRNA vaccine for cancer? https://www.dw.com/en/will-there-be-an-mrna-vaccine-for-cancer/a-58614316.
21.Carter, D. Can mRNA vaccines be used in cancer care? https://www.mdanderson.org/cancerwise/can-mrna-vaccines-like-those-used-for-covid-19-be-used-in-cancer-care.h00-159457689.html.
22.Heine, A.; Juranek, S.; Brossart, P. Clinical and Immunological Effects of MRNA Vaccines in Malignant Diseases. Mol. Cancer 2021, 20 (1), 52
23.Fisusi, F. A.; Akala, E. O. Drug Combinations in Breast Cancer Therapy. Pharm. Nanotechnol. 2019, 7 (1), 3–23.
24.Zeng, C.; Zhang, C.; Walker, P. G.; Dong, Y. Formulation and Delivery Technologies for MRNA Vaccines. Curr. Top. Microbiol. Immunol. 2020. https://doi.org/10.1007/82_2020_217.Pardi, N.; Hogan, M. J.; Porter, F. W.; Weissman, D. MRNA Vaccines — a New Era in Vaccinology. Nat. Rev. Drug Discov. 2018, 17 (4), 261–279.
25.Miao, L.; Zhang, Y.; Huang, L. MRNA Vaccine for Cancer Immunotherapy. Mol. Cancer 2021, 20 (1), 41.
26.Tuesday, J. 6. Can mrna vaccines fight pancreatic cancer? MSK clinical researchers are trying to find out. https://www.mskcc.org/news/can-mrna-vaccines-fight-pancreatic-cancer-msk-clinical-researchers-are-trying-find-out.
27.Kingsley, T. MRNA Cancer Therapy Now in Human Trials after Shrinking Mouse Tumours. Independa 2021.
28.Cancer Genome Atlas Network. Comprehensive Molecular Portraits of Human Breast Tumours. Nature 2012, 490 (7418), 61–70.
29.PIK3CA gene https://medlineplus.gov/genetics/gene/pik3ca/.
30.Petrau, C.; Cornic, M.; Bertrand, P.; Maingonnat, C.; Marchand, V.; Picquenot, J.-M.; Jardin, F.; Clatot, F. CD70: A Potential Target in Breast Cancer? J. Cancer 2014, 5 (9), 761–764.
31.Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Long-Term Outcomes for Neoadjuvant versus Adjuvant Chemotherapy in Early Breast Cancer: Meta-Analysis of Individual Patient Data from Ten Randomised Trials. Lancet Oncol. 2018, 19 (1), 27–39.
32.Coleman, W. B. Breast Cancer Personalized Medicine: Challenges and Opportunities. Am. J. Pathol. 2013, 183 (4), 1036–1037
33.Fang, L.; Barekati, Z.; Zhang, B.; Liu, Z.; Zhong, X. Targeted Therapy in Breast Cancer: What's New? Swiss Med. Wkly 2011, 141, w13231.
34.Gampenrieder, S. P.; Rinnerthaler, G.; Greil, R. Neoadjuvant Chemotherapy and Targeted Therapy in Breast Cancer: Past, Present, and Future. J. Oncol. 2013, 2013, 732047.
35.Hernando, C.; Ortega-Morillo, B.; Tapia, M.; Moragón, S.; Martínez, M. T.; Eroles, P.; Garrido-Cano, I.; Adam-Artigues, A.; Lluch, A.; Bermejo, B.; Cejalvo, J. M. Oral Selective Estrogen Receptor Degraders (SERDs) as a Novel Breast Cancer Therapy: Present and Future from a Clinical Perspective. Int. J. Mol. Sci. 2021, 22 (15), 7812
36.Ju, J.; Zhu, A.-J.; Yuan, P. Progress in Targeted Therapy for Breast Cancer. Chronic Dis. Transl. Med. 2018, 4 (3), 164–175..
37.Kaczmarek, J. C.; Kowalski, P. S.; Anderson, D. G. Advances in the Delivery of RNA Therapeutics: From Concept to Clinical Reality. Genome Med. 2017, 9 (1). https://doi.org/10.1186/s13073-017-0450-0.
38.Schütz, F.; Domschke, C.; Schneeweiss, A. Targeted Therapy of HER2-Negative Breast Cancer. Oncol. Res. Treat. 2016, 39 (3), 118–121.
39.Winters, S.; Martin, C.; Murphy, D.; Shokar, N. K. Breast Cancer Epidemiology, Prevention, and Screening. Prog. Mol. Biol. Transl. Sci. 2017, 151, 1–32.
40.S.; Aksoy, G. Identification of Risk Factors for Breast Cancer for Women in Istanbul. Çelik 2007.29.PIK3CA gene https://medlineplus.gov/genetics/gene/pik3ca/.
41.Franceschini, G.; Martin Sanchez, A.; Di Leone, A.; Magno, S.; Moschella, F.; Accetta, C.; Masetti, R. New Trends in Breast Cancer Surgery: A Therapeutic Approach Increasingly Efficacy and Respectful of the Patient. G. Chir. 2015, 36 (4), 145–152.
42. Garay, J. P.; Park, B. H. Androgen Receptor as a Targeted Therapy for Breast Cancer. Am. J. Cancer Res. 2012, 2 (4), 434–445.
43.Vergauwen, L.; Cherradi, Y. Scalable Purification of Plasmid DNA: Strategies and Considera-tions for Vaccine and Gene

Therapy Manufacturing, 2021. https://doi.org/10.1021/scimeetings.1c01214.
44. U.S. Breast Cancer Statistics https://www.breastcancer.org/symptoms/understand_bc/statistics.
45. Dai, X.; Xiang, L.; Li, T.; Bai, Z. Cancer Hallmarks, Biomarkers and Breast Cancer Molecular Subtypes. J. Cancer 2016, 7 (10), 1281–1294.
46. Yersal, O.; Barutca, S. Biological Subtypes of Breast Cancer: Prognostic and Therapeutic Implications. World J. Clin. Oncol. 2014, 5 (3), 412–424.

Author

Alexandra Giuliani is a junior at the International School of Monaco. She enjoys studying molecular biology and looks forward to researching biotechnology in college.

DOI: 10.36838/v4i4.10

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Potential of CRISPR-Cas System Treatment to Eradicate the COVID-19 Pandemic Caused by SARS-COV-2

İpek Ercan
Uskudar American Academy - Fenerbahçe Neighborhood. Tevfikpaşa St. Sinanoğlu Apt. 11/9 Kadıköy, Istanbul, 34627, Turkey

ABSTRACT: Over the past two years, scientists' efforts to develop vaccines and antiviral treatments against the COVID-19 pandemic have failed due to the lack of a proven therapy and the inefficacy of currently available vaccines against the virus' mutant forms. Therefore, the COVID-19 pandemic has forced scientists to develop highly reliable diagnostic and therapeutic procedures to quickly prevent the infection's spread. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR-associated protein (Cas) system is an RNA-directed adaptive immune system used as a gene-editing technology. Recent CRISPR application studies have shown that CRISPR effectors may limit the replication of double-stranded DNA or single-stranded RNA (ssRNA) viruses in mammalians. Because SARS-COV-2 contains a single-stranded, positive-sense RNA genome and structure, CRISPR-Cas systems are suitable for therapeutic use. Similarly, many clinical studies have investigated the CRISPR-Cas systems' potential for diagnosing SARS-COV-2 and have proven their effectiveness. CRISPR-Cas systems' success in quickly diagnosing SARS-COV-2 serves as a model for applying CRISPR-Cas systems to treat the Covid-19 pandemic. This review aims to investigate the uses of CRISPR-Cas-based gene-editing systems (CRISPR-Cas9, CRISPR-Cas12, CRISPR-Cas13) when treating patients during the COVID-19 pandemic, to analyze related articles, and to examine this novel technology's limits and possibilities.

KEYWORDS: Antiviral strategy; Covid-19 pandemic; CRISPR-Cas systems; Gene editing; SARS-COV-2.

Introduction

Identifying SARS-COV-2 positive individuals and treating them is essential to control the COVID-19 outbreak.[1] While reverse transcription-polymerase chain reaction (RT-PCR) is currently the gold standard for molecular detection, its drawbacks include expensive equipment costs, staff training requirements, the risk for false-positive/negative results, and long processing times. CRISPR-Cas systems were determined to successfully and quickly diagnose SARS-COV-2 cases. Thus, they were used for diagnoses in the COVID-19 outbreak.[2] To exemplify, Sherlock Biosciences' Sherlock™ CRISPR SARS-COV-2 kits were licensed by the Food and Drug Administration (FDA) for emergency usage to alleviate COVID-19 detection bottlenecks in the United States.[3] CRISPR-Cas systems' efficacy in quickly diagnosing SARS-COV-2 showcases its potential success in treating the COVID-19 pandemic. Similarly, CRISPR-Cas systems may be used therapeutically since CRISPR effectors limit DNA and RNA structures' replication, such as SARS-COV-2. For example, the CRISPR-Cas13-based approach for COVID-19 treatment successfully destroyed SARS-COV-2 sequences and a live IAV genome. The system is called Prophylactic Antiviral CRISPR in human cells (PAC-MAN) to suppress SARS-COV-2.[4] This review provides in-depth examinations of the examples above of CRISPR-Cas systems' usages to analyze how and to what extent they help limit and treat the COVID-19 pandemic. To strengthen one's understanding of these examinations, the review first discusses the COVID-19 pandemic and the virus and the general history, mechanism, and technicalities of CRISPR-Cas systems.

Discussion

SARS-COV-2 and COVID-19 Pandemic:

While (severe acute respiratory syndrome coronavirus) SARS-COV-2 was first identified as an atypical respiratory tract infection agent in Wuhan, a city in the Hubei province of China, in December 2019, the World Health Organization declared a SARS-COV-2-related pandemic on March 11, 2020, just one year later. Unfortunately, the Covid-19 epidemic remains significant as a worldwide hazard. Many countries' healthcare systems have been strained to breaking point, with some collapsing under the weight of the pandemic, and meticulous precautionary efforts to slow virus transmissions, such as quarantine, travel restrictions, and social isolation, have had catastrophic consequences for the global economy and society.[5]

SARS-COV-2 is a spherical, enclosed coronavirus with a single-stranded, positive-sense RNA genome encased in an extracellular membrane comprising spike glycoproteins.[6] It is a positive-sense ssRNA that primarily infects the upper and lower respiratory tracts. Furthermore, it causes illness by direct cytotoxicity and activation of host cytokine-mediated inflammation.[7] The infection process starts with the virus entering the cell by endocytosis or direct fusion of the viral envelope with the host membrane. Then, the virus uncoats and releases RNA into the cytoplasm.[8]

While serological tests for COVID-19 as an immunological assay can mainly detect viral antigens in respiratory secretions or antibodies in the blood, molecular methods for COVID-19 are based on SARS-COV-2 RNA detection in nasopharyngeal samples. Since droplets primarily transmit the disease, COVID-19 may be seen in asymptomatic

DOI: 10.36838/v4i4.11

people (30%). After an incubation period of 4–14 days, most people have mild to severe symptoms (55%) with anorexia, myalgias, anosmia, dysgeusia (dysfunction of the sense of taste), a sore throat, a headache, or rhinorrhea (runny nose). Around day five, nearly 30% of individuals experience severe dyspnea (shortness of breath) and fever. Patients with severe symptoms often deteriorate in the second week and need hospitalization for hypoxemia and bilateral pneumonia (75%). Most hospitalized patients receive conventional treatment, but 20% may rapidly worsen with severe complications, including acute respiratory distress syndrome, acute cardiac injury, acute kidney injury, and septic shock.[10]

At present, there is not a single specific antiviral therapy for COVID-19, and symptomatic supportive care remains the mainstay of treatment.[10,11] Nonetheless, several drugs are being evaluated for their efficacy as antiviral agents and are suggested for use in the National Institutes of Health COVID-19 therapy recommendations. Convalescent plasma treatment is an exemplary example of a modern therapeutic technique.[12]

History of the CRISPR-Cas system:

In 1987, during a study regarding *Escherichia coli*, Y. Ishino and colleagues discovered an unusual genetic structure composed of alternating repeat and non-repeat DNA sequences.[13] In 2002, Jansen and colleagues revealed that prokaryotes possess multiple distinguishing features in their chromosomes, with repetitive "spacer" sequences placed on each side of distinct ones. They developed the abbreviation CRISPR for "Clustered Regularly Interspaced Short Palindromic Repeats," referring to the results of their study. The researchers then discovered that numerous clusters of hallmark CRISPR-associated Cas genes were highly preserved and often next to repetitive elements.[14] The finding of sequence similarities between the spacer sections of CRISPRs and those of bacteriophages, archaeal viruses, and plasmids in the early 2000s provided insight on CRISPR's immune system role. The CRISPR arrays' specific sequences were discovered in 2005 by three different research teams (Mojica and colleagues, Bolotin and colleagues, and Pourcel and colleagues) through the systematic examination of the spacer sequences between individual direct repeats.[15, 16, 17]

Many researchers proposed that CRISPR spacers operate as RNA-like guides to degrade viral transcripts or spacer-matching CRISPR spacers drive Cas enzymes to cleave viral DNA at spacer-matching areas.[16,18] In a subsequent study published by Rodolphe Barrangou *et al*. of Danisco, a yogurt company, researchers employed *Streptococcus thermophilus* bacteria as models. They noticed that they were implanted by inserting new pieces into areas of bacteria that survive a viral assault. Additionally, they determined that the DNA sequence of these spacers is similar to that of segments of the viral genome. Moreover, the researchers altered it by eliminating spacers and inserting new viral DNA sequences. They could modify the bacterium's resistance to a particular viral attack. Consequently, the researchers established that CRISPR is crucial for bacterial immunity regulation.[19] Individual spacers exist to guide Cas nuclease activity.[20] In 2013, Cas was confirmed as an RNA-guided endonuclease.[21] Makarova et al. discovered that one or more operons having a cluster of Cas genes producing the system's effector enzymes are located near the CRISPR array.[22] Then, CRISPR-Cas was applied in eukaryotic cells. In 2014, genome-wide functional screening using Cas studies was done.[23] In 2016, Cong and Hsu successfully made in vivo genome editing via homolog independent targeted integration.[21, 24]

The function of the CRISPR-Cas system:

CRISPR-Cas is an RNA-directed adaptive immune system that occurs naturally in about 48% of bacteria and 95% of archaea. The system is triggered by the invasion of foreign genetic material.[9, 25] It generates an immune response in three stages: adaptation, pre-crRNA expression/processing, and interference.[26] Firstly, during the adaptation step, short and direct repeats separated by spacers constitute CRISPR sequences. Then, a foreign DNA termed protospacer is cleaved and incorporated into the CRISPR array. As a result, the integrated pieces function as new spacers.[26] The second step is the expression, during which the CRISPR array is transcribed to generate precursor CRISPR-derived RNA (pre-crRNA), which is subsequently matured to develop CRISPR-derived RNA (crRNA).[27] This is proceeded by the interference phase, which attaches the mature crRNA to the processing complex and serves as a guide RNA, recognizing identical sequences in the invading viral RNA. Viral RNA is subsequently cleaved and inactivated by one of the Cas proteins.[28]

CRISPR-Cas systems' capacity to produce a double-strand break (DSB) at a particular genomic locus ensures their use for genome editing. The occurrence of DSBs relies entirely on the host cell's DNA repair mechanism to fix the lesion created by these systems. The repair methods may be homology-directed repair (HDR) or non-homologous end joining (NHEJ). HDR repairs the DSB by using a homologous template DNA to the break site, such as an unbroken sister chromatid or homologous chromosome. Then, it delivers exogenous DNA Templates to the host genome to affect a user-defined modification. In contrast, NHEJ relies on the direct combining of the DSB's broken ends, rendering NHEJ the more error-prone of the two mechanisms. NHEJ may be used to damage genes, while HDR enables the introduction of new genetic data or the direct repair of a single locus's sequence. The NHEJ is a repair mechanism in which DSB junctions undergo insertions and deletions. Although the HDR process needs the presence of a homologous DNA template, it is very accurate in repairing DSBs, and exogenous homologous sequences may be used to direct genome editing.[29]

Classification of the CRISPR-Cas systems:

Each with its unique composition and method of action, CRISPR-Cas systems have six CRISPR-Cas types in total, and at least 29 subtypes 6 -8, and both the type and subtype lists are constantly expanding.[29] The CRISPR-Cas system is classified into two distinct groups, depending on the arrangement of its effectors.[9] The Class 1 system cleaves the target genome sequence using crRNA and a multi-effector complex. In the class I system, the ribonucleoprotein (RNP) complex has several protein subunits and crRNA. In contrast, the Class

DOI: 10.36838/v4i4.11

2 system cleaves the target genome sequence using a single multidomain Cas protein and crRNA for interference. The Class II system contains just one protein and crRNA to target invasive viral RNAs.[30] Class 2 CRISPR-Cas systems are found exclusively in bacteria and combine a crRNA with a Cas protein to form a ribonucleoprotein complex. The crRNA is programmed to recognize a PAM sequence adjacent to the target DNA.[31] These multidomain effector proteins interfere through complementarity between the crRNA and the target sequence.[30] Each of the two classes is further subdivided into three subtypes, Type I, III, and IV in Class I and Type II, V, and VI in Class II.[18]

CRISPR-associated (Cas) genes are a group of genes with varied orientations and sequences that code encoding Cas proteins. They are critical for acquiring and destroying foreign sequences. A total of 93 distinct Cas genes have been recognized to date. Based on sequence similarity, these genes were categorized into 35 families. This review focused on Cas proteins: Cas9, Cas12, Cas13, classified as type II, type V, and type VI enzymes.[31]

The most widely characterized CRISPR-Cas system is the type II subtype II-A found in *Streptococcus pyogenes* (Sp), which uses SpCas9. Cas9 was the first Cas-protein engineered for use in gene editing[32], and it is the most efficient genome-editing machinery for targeting double-stranded DNA (dsDNA) or single-stranded DNA (ssDNA). Cas9 uses trans-activating CRISPR RNA (tracrRNA) and RNase III to process the pre-crRNA in the type II CRISPR-Cas system.[33] In comparison, the Cas12 and Cas13 proteins in the type V and VI systems process the pre-crRNA directly. While Cas12a and Cas12b proteins break dsDNA upon identification by matured crRNA, Cas13 protein cleaves ssRNA. Additionally, Cas13 is not dependent on the presence of a PAM in the target RNA, while Cas12 needs a PAM in the dsDNA target but not in the ssDNA target.

CRISPR-Cas systems are being used as an antiviral, targeting genes in human and viral genomes. Studies have effectively targeted viral and host genes involved in viral entry, replication, and persistence.[34] The invention, advancement, and ease of using Type II CRISPR-Cas9 systems have expedited their adoption and implementation across various applications.

CRISPR-Cas systems versus pathogenic viruses:

Class 2 systems are used in various fields, including genome editing, which can be accomplished using a type II or type V effector enzyme to evaluate potential developing therapies and target ssDNA/RNA and for viral disease detection and treatment.[35]

Current Cas9 applications have shown that CRISPR effectors may limit replication of dsDNA or single-stranded RNA (ssRNA) viruses with DNA intermediates in mammalian cells.[7, 23, 36-39] Liu *et al*. found from their research that HBV-specific gRNA/Cas9 systems inhibit HBV replication in different genotypes. HBV templates are cleaved and cleared by HBV-specific gRNA/Cas9 systems, and so they proved the inhibition of HBV *in vivo* by the gRNA/Cas9 system induces clear.[7] Then, the work by Ophinni *et al*. revealed that CRISPR-Cas9 might target the HIV-1 proviral genome and suppress replication in latency models.[36] Later, Ramanan *et al*. demonstrated that the CRISPR-Cas9 system could selectively target and cleave conserved areas of the HBV genome, resulting in potent inhibition of viral gene expression and replication.[37] Lastly, Roehm *et al*. revealed that Cas9 and gRNA delivery by Lentivirus reduces HSV-1 infection and protects cells against infection.[38]

Cas9 and other DNA-targeting effectors defend against invading DNA bacteriophages. However, around two-thirds of viruses capable of infecting humans have ssRNA genomes, and just 2.5 percent of those viruses include DNA intermediates that could be targeted with Cas9.[39, 40] Recently identified Cas9 orthologs that target RNA and DNA are less likely to satisfy this demand, since they have poor RNA cleavage performance and may produce off-target effects on cellular DNA.[41]

It was then discovered that Cas 12 and Cas13 effectors, which are associated with type III and VI CRISPR systems, may contribute to the defense of mammalian cells against DNA and RNA viruses.[42] CRISPR-Cas12 is often referred to as CRISPR-associated endonuclease. From Prevotella and Francisella 1 (cpf1), it is a type V-associated nuclease belonging to the Class 2 family.[43, 44] The CRISPR-Cas12 system is an efficient method to generate staggered cuts in both ssDNA and dsDNA, and it utilizes only the crRNA to make staggered cuts at the targeted region. In rodent models of influenza and SARS-COV-2 infections in 2021, Cas13a was shown to attenuate the severity of influenza and SARS-COV-2 *in vivo*.[45] Cas13a was supplied in a synthetic mRNA that allows for repeat dosing since the expression is temporary. It was the first time that Cas13a has been tested as a therapy for influenza infection after it had been administered in vivo. There were no *in vivo* tests for treating SARS-COV-2 conditions, since Cas13a was administered before infection.

Applications of the CRISPR-Cas systems in the diagnosis of SARS-COV-2:

RNA-guided CRISPR-Cas nuclease-based nucleic acid detection has recently shown tremendous promise for developing next-generation molecular diagnostics technologies because of its high sensitivity, reliability, and specificity. Due to the challenges associated with heterologous expression of multiple cascade complexes, the class 2 CRISPR-Cas system is more convenient and quick in responding to disasters such as COVID-19 than the class 1 system. For instance, since single-gene testing may produce false-negative results, Xiong *et al*. offered dual-gene testing, a widely adopted strategy by clinical-approved RT-qPCR diagnostic kits. They improved the CRISPR-Cas9-mediated triple-line LFA (TL-LFA) design combined with multiplex reverse transcription-recombinase polymerase amplification (RT-RPA). As a result, an examination of 64 clinical samples taken from the nose revealed that the CRISPR-Cas9-mediated T-LFA approach is as analytically specific and sensitive as the gold-standard RT-qPCR procedure.[46]

For rapid, specific, and sensitive detection of SARS-COV-2, Ali *et al*. developed Bio-SCAN (biotin-coupled

specific CRISPR-based assay for nucleic acid detection) for pathogen detection. Ali *et al*. tested Bio-SCAN's ability to identify SARS-COV-2 variations by focusing on the spike polypeptide alterations that increase infectivity and disseminate COVID-19 worldwide. They purposed to demonstrate that one could use the Bio-SCAN platform to forecast the emerging trends and prevalence of specific variant(s) in a particular sample collection from a population group or region by designing five sgRNAs with a specific focus on the PAM or seed region of the target sequence of the prevalent variants. Bio-SCAN diagnosed SARS-COV-2 with acceptable accuracy, demonstrating its use for routine examination in non-laboratory POC situations. Bio-SCAN can accurately identify individual variations, enabling trend prediction or identification of the most common SARS-COV-2 variant(s) during screening tests.[47]

Cas13d is one of the best subtypes because of its high precision, strong knockdown efficiency, and short coding sequence, making it easy to spread. The unique aspect of the Cas13d enzyme is that its cleavage does not depend on PAM-like sequences. This makes it easier to target constantly changing viruses because the rapid development of crRNAs makes it easier to target them. To illustrate, Nguyen *et al*. used CRISPR-Cas13d technology to build a flexible and efficient technique for targeting RNA in the laboratory. To reduce the capacity of production SARS-COV-2, they preferred to use the CRISPR-Cas13d system because of its flexibility in making guide RNAs and the unnecessary use of PAM. Also, they offered that adeno-associated virus (AAV) could be used to deliver the Cas13d effector to people who have been infected with SARS-COV-2. Because of the small size of the Cas13d effector, it can be used for an "all-in-one" AAV delivery with a gRNA array. In addition, AAV has serotypes specific to the lung, which is the main organ infected by SARS-COV-2. This means that the CRISPR system can be delivered to the right place. A similar strategy can also be used to fight other RNA viruses. Nguyen *et al*. developed the system with engineered crRNAs and optimized conditions to detect various clinically relevant nucleic acid targets, including human immunodeficiency virus, hepatitis virus C, and SARS-COV-2, with high sensitivity.[48]

The review had previously given examples of the potential of CRISPR-Cas systems, including Cas9, Cas12, and Cas13 proteins, for diagnosing SARS-COV-2 infection in recently completed studies. Following the Cas9 protein, the Cas12 and Cas13 proteins focus on a viral illness diagnosis. Cas12 cleaves both ssDNA and dsDNA, while Cas13 cleaves only ssRNA; hence, both proteins are critical in diagnosing SARS-COV-2.[48]

As an example of Cas12 utilization, Brandsma *et al*. developed a diagnostic platform based on the Cas12a collateral cleavage of a reporter nucleic acid composed of an ssDNA called DETECTR (DNA endonuclease-targeted CRISPR trans reporter) to recognize and detect SARS-COV-2 gRNA. Comparing DETECTR with qRT-PCR to diagnose SARS-COV-2, they found that both techniques are equally sensitive in detecting SARS-COV-2, and DETECTR was found 100% specific for SARS-COV-2 relative to other human coronaviruses.[49] HOLMES, a one-hour low-cost, highly efficient system, is another Cas12a-based detector for precise target DNA detection.[44] Another example would be the All-In-One Dual CRISPR-Cas12 (AIOD-CRISPR) assay, introduced by Ding *et al*. as an ultra-sensitive and faster assay to detect the target nucleic acid. These AIOD-CRISPR assay mixtures are incubated at 37 ˚C to prevent contamination, simplifying the detection procedure. Utilizing AIOD-CRISPR assay, they caught the RNA of HIV and SARS-COV-2 within 60 minutes.[50]

As opposed to Cas9, Cas13 causes loss-of-function phenotypes without causing chromosomal deletion of the targeted gene. So, it meets the demand for quick gRNA creation to target virus strains that change and elude traditional drugs. Using Cas9, which cleaves DNA, has several disadvantages over Cas13, which cleaves RNA. Off-target cleavage and mismatches may cause unintended mutations and malignancy. Moreover, Cas9 often cleaves DNA at the proper site, but the cellular repair process fails. Thus, CRISPR-Cas13 may be employed directly in tissues damaged by SARS-COV-2 as a possible treatment strategy.[51]

An example of Cas13 usage is the SHERLOCK technology, which utilized a quenched fluorescent ssRNA reporter. It is the first platform built on CRISPR-Cas13 systems, enabling bio-sensing with sensitivity for DNA and RNA virus detection with single-base distinction. This Cas13a based system also has been used for the robust detection of Zika and dengue viruses.[52] As a result, SHERLOCK for SARS-COV-2 detection is significantly faster than qRT-PCR and has a sensitivity of 93.1 percent. Gootenberg *et al*. developed SHERLOCK further to create STOP (SHERLOCK Testing in One Pot) for specific point-of-care (POC) diagnosis of COVID-19. They demonstrated significant sensitivity and specificity in detecting Zika and Dengue viruses using SHERLOCK. Additionally, SHERLOCK was used to detect the Ebola and Lassa viruses. When the COVID-19 epidemic began, Feng Zhang's team recreated SHERLOCK to make SARS-COV-2 detection possible. They modified different primers and guided RNA to deliberately target the open reading frame 1ab (ORF1ab) and spike (S) genes. Clinical verifications indicated that the SHERLOCK test result is equal to RT-qPCR. They demonstrated that SHERLOCK is a flexible, reliable method for detecting RNA and DNA that may be used for quick diagnoses, such as infectious disease applications and sensitivity genotyping.[51] Sherlock Biosciences was recently licensed by the Food and Drug Administration (FDA) for emergency usage to alleviate COVID-19 detection bottlenecks in the United States.[3]

Because SARS-COV-2 viral load varies during the day and illness phases, a quantitative reverse transcriptase PCR (qRT–PCR) diagnostic approach may be negative when the viral load is low, necessitating a more reliable test. Rauch *et al*. devised a CREST method (Cas13-based, rugged, equitable, scalable testing) for detecting SARS-COV-2. They confirmed that CREST is equivalent sensitivity to the gold standard reverse transcription-quantitative PCR (RT-qPCR) method.[54] The

DOI: 10.36838/v4i4.11

accuracy in detecting target templates through the Cas-sg/crRNA complex makes CRISPR-Cas technologies an outstanding alternative to PCR-based approaches.[53]

Cas13d is significant among all subtypes because of its high efficiency and strong knockdown efficiency, and its simplicity of viral administration due to the effector domain's short coding sequence. Recent research focused on repurposing the RNA-guided RNA endonuclease activity of Cas13d in mammalian cells against SARS-COV-2 and live influenza A virus (IAV). PAC-MAN, which stands for "prophylactic antiviral CRISPR in huMAN cells," is the first Cas13 antiviral strategy to fight SARS-COV-2. Abbott *et al*. presented a CRISPR-Cas13-based approach for COVID-19 treatment called Prophylactic Antiviral CRISPR in huMAN cells (PAC-MAN) to suppress SARS-COV-2.[4] In this study, Cas13d is guided by pan-coronavirus crRNAs to destroy the viral genome and suppress gene expression. As there was no access to live SARS-COV-2 strains at the time of this investigation (April 2020), the researchers resorted to synthesized fragments of SARS-COV-2 (besides using live H1N1 IAV). Abbott and colleagues used bioinformatic screens to identify highly conserved regions across the viral genomes. They found a group of six crRNAs that could target 91% of all coronaviruses and a group of 22 crRNAs that could target all coronaviruses with no mismatches. The PAC-MAN approach successfully destroyed SARS-COV-2 sequences and a live IAV genome. Cas13d may specifically target conserved genomic areas of SARS-COV-2. This way, CRISPR-Cas technology might target viral RNA for destruction and limit virus reproduction in host cells, restricting transmission of the virus.[4]

The CRISPR-Cas system has been used to regulate epigenetic modifications in eukaryotic systems to correct genetic errors and improve hereditary traits. Cas12 and Cas13 are CRISPR proteins that are efficient agents for diagnosing and combating ssRNA viruses. The development of COVID, RNA viruses have been a primary worldwide concern, and Cas13 has garnered considerable attention as an antiviral due to its ability to cut ssRNA, such as that seen in SARS-COV-2. Since ssRNA viruses account for the bulk of viruses capable of infecting people, a Cas13 antiviral system was created and confirmed in mammalian cells before COVID.[42] Freije *et al*. tested Cas13's action against three different ssRNA viruses in the laboratory [lymphocytic choriomeningitis virus (LCMV), influenza A virus (IAV), and vesicular stomatitis virus (VSV)]. They created Cas13-assisted restriction of viral expression and readout (CARVER) by combining this antiviral activity with Cas13-based diagnosis. The researchers identified Cas13 crRNA target sites in viral RNA. They then used cell culture models to create a series of antiviral crRNAs that could be multiplexed in a combinatorial fashion, demonstrating that the crRNA-directed Cas13 enzyme effectively inhibits viral RNA replication in mammalian cells.[42] The researchers demonstrated in their study that the CARVER system for quick diagnostics and antiviral drug development and CRISPR RNAs (crRNAs) targeting highly conserved regions of the viral genome were advantageous to avoid escape mutants. They noticed that this technique is remarkably adaptive to resist probable viral evolution since it employs multiplexing (targeting many loci) with pooled crRNAs.[42]

Conclusion

This study examined the potential efficacy and utility of CRISPR-Cas technology in limiting the COVID-19 pandemic. For this purpose, we evaluated the mechanism of CRISPR-CAS systems, functions of Cas9, Cas12, and Cas13, their applications in diagnosis and targeting SARS-COV-2, and the history of CRISPR technology and the pandemic. Cas13d was determined to be the most significant among all subtypes because of its high efficiency, strong knockdown efficiency, and ease of viral administration due to the effector domain's short coding sequence. Cas13 effectors have been successfully cut complementary target ssRNA, making it a more reliable and protected alternative to Cas9 since it produces loss-of-function phenotypes without triggering chromosomal loss of the targeted gene. Moreover, Cas13 does not need Cas9 to act on a PAM, making the CRISPR-Cas13d system adaptable in guiding RNA design. There have been reports of CRISPR-based antiviral agents utilizing Cas9 inducing mutations in target sites[23], but no crRNA target-site mutations were identified after Cas13 treatment.[42] These findings suggest that the Cas13 enzyme can be an efficient antiviral agent in treating SARS-COV-2 infection. As a result, CRISPR-Cas12 systems satisfy the need for fast gRNA generation to target various virus strains that mutate and evade conventional treatments. Although these results are promising, CRISPR-Cas systems are still not ready to be used for fighting acute viruses since various problems require solutions before antivirals are utilized in in-body work. There is a risk that genome editing therapies for viruses could change the human genome. The consensus opinion is that genome editing in somatic cells is acceptable if the goal is to treat horrible diseases, but gene editing is fairly disheartening (National Academy of Sciences, 2020). In this way, genome editing treatment methods for viral infectious diseases are likely to be less controversial because they only affect cells in the body.[55] Furthermore, the immunogenicity of Cas proteins has not yet been thoroughly explored, and further safety studies are required to determine how and when they are safely expressed in people.[56]

References

1. Pollard, C. A.; Morran, M. P.; Nestor-Kalinoski, A. L. The COVID-19 pandemic: A global health crisis. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7686876/ (accessed Jan 30, 2022).
2. Nouri, R.; Tang, Z.; Dong, M.; Liu, T.; Kshirsagar, A.; Guan, W. CRISPR-based detection of SARS-COV-2: A review from sample to result.
3. Satyanarayana, M. A COVID-19 diagnostic that uses CRISPR gets a nod from the FDA. https://cen.acs.org/analytical-chemistry/diagnostics/COVID-19-diagnostic-uses-CRISPR/98/web/2020/05 (accessed Jan 30, 2022).
4. Abbott, T. R.; Dhamdhere, G.; Liu, Y.; Lin, X.; Goudy, L.; Zeng, L.; Chemparathy, A.; Chmura, S.; Heaton, N. S.; Debs, R.; Pande, T.; Endy, D.; La Russa, M. F.; Lewis, D. B.; Qi, L. S. Development of CRISPR as an antiviral strategy to combat SARS-COV-2 and influenza. https://www.ncbi.nlm.nih.gov/pmc/articles/

PMC7189862/ (accessed Jan 29, 2022).
5. Aliouche, H. How to prevent pandemics. https://www.news-medical.net/health/How-to-Prevent-Pandemics.aspx (accessed Jan 29, 2022).
6. Pal, M.; Berhanu, G.; Desalegn, C.; Kandi, V. Severe acute respiratory syndrome coronavirus-2 (SARS-COV-2): An Update. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7182166/. (Accessed Jan 30, 2022).
7. Liu, X.; Hao, R.; Chen, S.; Guo, D.; Chen, Y. Inhibition of hepatitis B virus by the CRISPR/cas9 system via targeting the conserved regions of the viral genome. https://www.microbiologyresearch.org/content/journal/jgv/10.1099/vir.0.000159#tab2. (Accessed Jan 30, 2022).
8. Sharma, A.; Balda, S.; Apreja, M.; Kataria, K.; Capalash, N.; Sharma, P. Covid-19 diagnosis: Current and future techniques. https://www.sciencedirect.com/science/article/pii/S0141813021024144 (accessed Jan 30, 2022).
9. Safari, F.; Afarid, M.; Rastegari, B.; Borhani-Haghighi, A.; Barekati-Mowahed, M.; Behzad-Behbahani, A. CRISPR systems: Novel approaches for detection and combating COVID-19. https://www.sciencedirect.com/science/article/pii/S0168170220311898. (Accessed Jan 30, 2022).
10. Baj, J.; Karakuła-Juchnowicz, H.; Teresiński, G.; Buszewicz, G.; Ciesielka, M.; Sitarz, E.; Forma, A.; Karakuła, K.; Flieger, W.; Portincasa, P.; Maciejewski, R. Covid-19: Specific and non-specific clinical manifestations and symptoms: The current state of knowledge. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7356953/ (accessed Jan 29, 2022).
11. Chan, K. G.; Ang, G. Y.; Yu, C. Y.; Yean, C. Y. Harnessing CRISPR-Cas to combat COVID-19: From diagnostics to therapeutics. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8623262/. (accessed Jan 29, 2022).
12. Information on Covid-19 treatment, prevention and research. https://www.covid19treatmentguidelines.nih.gov/ (accessed Jan 30, 2022).
13. Ishino, Y.; Krupovic, M.; Forterre, P. History of CRISPR-Cas from encounter with a mysterious repeated sequence to genome editing technology. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5847661/ (accessed Jan 29, 2022).
14. Jansen, R.; Embden, J. D. A. van; Gaastra, W.; Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. https://onlinelibrary.wiley.com/doi/full/10.1046/j.1365-2958.2002.02839.x?sid=nlm%3Apubmed. (Accessed Jan 29, 2022).
15. Mojica, F. J. M.; Díez-Villaseñor, C.; García-Martínez, J.; Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. https://pubmed.ncbi.nlm.nih.gov/15791728/ (accessed Jan 30, 2022).
16. Bolotin, A.; Quinquis, B.; Sorokin, A.; Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. https://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.28048-0#tab2. (Accessed Jan 29, 2022).
17. Pourcel, C.; Salvignol, G.; Vergnaud, G. CRISPR elements in yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. https://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.27437-0#tab2. (Accessed Jan 30, 2022).
18. Makarova, K. S.; Wolf, Y. I.; Alkhnbashi, O. S.; Costa, F.; Shah, S. A.; Saunders, S. J.; Barrangou, R.; Brouns, S. J. J.; Charpentier, E.; Haft, D. H.; Horvath, P.; Moineau, S.; Mojica, F. J. M.; Terns, R. M.; Terns, M. P.; White, M. F.; Yakunin, A. F.; Garrett, R. A.; van der Oost, J.; Backofen, R.; Koonin, E. V. An updated evolutionary classification of CRISPR-Cas Systems. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5426118/. (Accessed Jan 30, 2022).
19. Barrangou, R.; Fremaux, C.; Deveau, H.; Richards, M.; Boyaval, P.; Moineau, S.; Romero, D. A.; Horvath, P. CRISPR provides acquired resistance against viruses in prokaryotes. https://pubmed.ncbi.nlm.nih.gov/17379808/. (Accessed Jan 29, 2022).
20. Brouns, S. J. J.; Jore, M. M.; Lundgren, M.; Westra, E. R.; Slijkhuis, R. J. H.; Snijders, A. P. L.; Dickman, M. J.; Makarova, K. S.; Koonin, E. V.; van der Oost, J. Small CRISPR RNAS Guide antiviral defense in prokaryotes. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5898235/. (accessed Jan 29, 2022).
21. Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A.; Zhang, F. Multiplex Genome Engineering using CRISPR/Cas Systems. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3795411/ (accessed Jan 30, 2022).
22. Makarova, K. S.; Grishin, N. V.; Shabalina, S. A.; Wolf, Y. I.; Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic rnai, and hypothetical mechanisms of action.
23. Wang, L.; Zhou, J.; Wang, Q.; Wang, Y.; Kang, C. Rapid design and development of CRISPR-CAS13A targeting SARS-COV-2 spike protein. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7738867/ (accessed Jan 30, 2022).
24. Hsu, P. D.; Lander, E. S.; Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4343198/. (Accessed Jan 29, 2022).
25. Zhan, Y.; Li, X.-P.; Yin, J.-Y. Covid-19 one year later: A retrospect of CRISPR-cas system in combating covid-19. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8193275/. (Accessed Jan 30, 2022).
26. Amitai, G.; Rotem, S. Intracellular signaling in CRISPR-Cas Defense. https://pubmed.ncbi.nlm.nih.gov/28798118/ (accessed Jan 29, 2022).
27. Doudna, J. A.; Charpentier, E. The New Frontier of Genome Engineering with CRISPR-Cas9. https://www.science.org/doi/10.1126/science.1258096?url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub++0pubmed (accessed Jan 30, 2022).
28. Plagens, A.; Tjaden, B.; Hagemann, A.; Randau, L.; Hensel, R. Characterization of the CRISPR/cas subtype I-A system of the hyperthermophilic crenarchaeon Thermoproteus Tenax. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3347209/. (accessed Jan 30, 2022).
29. Nidhi, S.; Anand, U.; Oleksak, P.; Tripathi, P.; Lal, J. A.; Thomas, G.; Kuca, K.; Tripathi, V. Novel CRISPR-Cas Systems: An updated review of the current achievements, applications, and Future Research Perspectives.
30. Shmakov, S.; Smargon, A.; Scott, D.; Cox, D.; Pyzocha, N.; Yan, W.; Abudayyeh, O. O.; Gootenberg, J. S.; Makarova, K. S.; Wolf, Y. I.; Severinov, K.; Zhang, F.; Koonin, E. V. Diversity and evolution of class 2 CRISPR-Cas Systems.
31. Koonin, E. V.; Makarova, K. S. Origins and evolution of CRISPR-Cas Systems. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6452270/ (accessed Jan 30, 2022).
32. Paul, B.; Montoya, G. CRISPR-Cas12a: Functional overview and applications. https://www.sciencedirect.com/science/article/pii/S2319417019305050. (Accessed Jan 30, 2022).
33. Deltcheva, E.; Chylinski, K.; Sharma, C. M.; Gonzales, K.; Chao, Y.; Pirzada, Z. A.; Eckert, M. R.; Vogel, J.; Charpentier, E. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3070239/. (Accessed Jan 29, 2022).

DOI: 10.36838/v4i4.11

34.Koujah, L.; Shukla, D.; Naqvi, A. R. CRISPR-cas based target ing of host and viral genes as an antiviral strategy. https://www. ncbi.nlm.nih.gov/pmc/articles/PMC6783345/. (Accessed Jan 29, 2022).
35.Hillary, V. E.; Ignacimuthu, S.; Ceasar, S. A. Potential of CRISPR/ CAS system in the diagnosis of COVID-19 infection. https:// www.ncbi.nlm.nih.gov/pmc/articles/PMC8607542/. (Accessed Jan 29, 2022).
36.Ophinni, Y.; Inoue, M.; Kotaki, T.; Kameoka, M. CRISPR/cas9 system targeting regulatory genes of HIV-1 inhibits viral replica tion in infected T-cell cultures. https://www.ncbi.nlm.nih.gov/ pmc/articles/PMC5958087/ (accessed Jan 30, 2022).
37.Ramanan, V.; Shlomai, A.; Cox, D. B. T.; Schwartz, R. E.; Michai lidis, E.; Bhatta, A.; Scott, D. A.; Zhang, F.; Rice, C. M.; Bhatia, S. N. CRISPR/Cas9 cleavage of viral DNA efficiently suppresses hepatitis B virus. https://www.ncbi.nlm.nih.gov/pmc/articles/ PMC4649911/. (Accessed Jan 30, 2022).
38.Roehm, P. C.; Shekarabi, M.; Wollebo, H. S.; Bellizzi, A.; He, L.; Salkind, J.; Khalili, K. Inhibition of HSV-1 replication by gene editing strategy. https://www.ncbi.nlm.nih.gov/pmc/articles/ PMC4827394/. (Accessed Jan 30, 2022).
39.Yin, H.; Song, C.-Q.; Dorkin, J. R.; Zhu, L. J.; Li, Y.; Wu, Q.; Park, A.; Yang, J.; Suresh, S.; Bizhanova, A.; Gupta, A.; Bolukbasi, M. F.; Walsh, S.; Bogorad, R. L.; Gao, G.; Weng, Z.; Dong, Y.; Ko teliansky, V.; Wolfe, S. A.; Langer, R.; Xue, W.; Anderson, D. G. Therapeutic genome editing by combined viral and non-viral de livery of CRISPR system components in vivo. https://www.ncbi. nlm.nih.gov/pmc/articles/PMC5423356/. (Accessed Jan 30, 2022).
40.Woolhouse, M. E. J.; Adair, K. The diversity of human RNA virus es. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5831953/ (accessed Jan 30, 2022).
41.Strutt, S. C.; Torrez, R. M.; Kaya, E.; Negrete, O. A.; Doudna, J. A. RNA-dependent RNA targeting by CRISPR-Cas9. https:// www.ncbi.nlm.nih.gov/pmc/articles/PMC5796797/. (Accessed Jan 30, 2022).
42.Freije, C. A.; Myhrvold, C.; Boehm, C. K.; Lin, A. E.; Welch, N. L.; Carter, A.; Metsky, H. C.; Luo, C. Y.; Abudayyeh, O. O.; Gootenberg, J. S.; Yozwiak, N. L.; Zhang, F.; Sabeti, P. C. Pro grammable inhibition and detection of RNA viruses using CAS13. https://www.ncbi.nlm.nih.gov/sites/ppmc/articles/ PMC7422627/ (accessed Jan 29, 2022).
43.Lewis, K. M.; Ke, A. Building the Class 2 CRISPR-Cas Arsenal. https://www.cell.com/molecular-cell/fulltext/S1097-2765(17)30050-3?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS1097276517300503%3Fshowall%3Dtrue. (Accessed Jan 30, 2022).
44.Li, S.-Y.; Cheng, Q.-X.; Wang, J.-M.; Li, X.-Y.; Zhang, Z.-L.; Gao, S.; Cao, R.-B.; Zhao, G.-P.; Wang, J. CRISPR-CAS12A-assisted nucleic acid detection. https://www.nature.com/articles/s41421-018-0028-z. (Accessed Jan 30, 2022).
45.Blanchard, E. L.; Vanover, D.; Bawage, S. S.; Tiwari, P. M.; Rotolo, L.; Beyersdorf, J.; Peck, H. E.; Bruno, N. C.; Hincapie, R.; Michel, F.; Murray, J.; Sadhwani, H.; Vanderheyden, B.; Finn, M. G.; Brinton, M. A.; Lafontaine, E. R.; Hogan, R. J.; Zurla, C.; Santangelo, P. J. Treatment of influenza and SARS-COV-2 in fections via mrna-encoded CAS13A in rodents. https://www.nature.com/articles/s41587-021-00822-w (accessed Jan 29, 2022).
46.Xiong, E.; Jiang, L.; Tian, T.; Hu, M.; Yue, H.; Huang, M.; Lin, W.; Jiang, Y.; Zhu, D.; Zhou, X. Simultaneous dual-gene diagnosis of SARS-COV-2 based on CRISPR/Cas9-mediated lateral flow assay. https://onlinelibrary.wiley.com/doi/full/10.1002/ange.202014506 (accessed Jan 30, 2022).
47.Ali, Z.; Sánchez, E.; Tehseen, M.; Mahas, A.; Marsic, T.; Aman, R.; Sivakrishna Rao, G.; Alhamlan, F. S.; Alsanea, M. S.; Al-Qahtani, A. A.; Hamdan, S.; Mahfouz, M. Bio-SCAN: A CRISPR/dcas9-based lateral flow assay for rapid, specific, and sensitive detection of SARS-COV-2. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8751644/ (accessed Jan 30, 2022).
48.Nguyen, T. M.; Zhang, Y.; Pandolfi, P. P. Virus against virus: A potential treatment for 2019-ncov (SARS-COV-2) and other RNA viruses.
49.Brandsma, E.; Verhagen, H. J. M. P.; van de Laar, T. J. W.; Claas, E. C. J.; Cornelissen, M.; van den Akker, E. Rapid, sensitive, and specific severe acute respiratory syndrome coronavirus 2 detection: A Multicenter comparison between standard quantitative reverse-transcriptase polymerase chain reaction and CRISPR-based DETECTR. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7665660/. (Accessed Jan 29, 2022).
50.Ding, X.; Yin, K.; Li, Z.; Liu, C. All-in-one dual CRISPR-CAS12A (AIOD-CRISPR) assay: A case for rapid, ultrasensitive and visual detection of novel coronavirus SARS-COV-2 and HIV virus. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7239053/. (Accessed Jan 29, 2022).
51.Lotfi, M.; Rezaei, N. CRISPR/cas13: A potential therapeutic option of covid-19. https://www.sciencedirect.com/science/article/pii/S0753332220309318. (Accessed Jan 30, 2022).
52.Gootenberg, J. S.; Abudayyeh, O. O.; Lee, J. W.; Essletzbichler, P.; Dy, A. J.; Joung, J.; Verdine, V.; Donghia, N.; Daringer, N. M.; Freije, C. A.; Myhrvold, C.; Bhattacharyya, R. P.; Livny, J.; Regev, A.; Koonin, E. V.; Hung, D. T.; Sabeti, P. C.; Collins, J. J.; Zhang, F. Nucleic acid detection with CRISPR-CAS13A/C2C2. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5526198/. (Accessed Jan 29, 2022).
53.Rauch, J. N.; Valois, E.; Solley, S. C.; Braig, F.; Lach, R. S.; Audouard, M.; Ponce-Rojas, J. C.; Costello, M. S.; Baxter, N. J.; Kosik, K. S.; Arias, C.; Acosta-Alvear, D.; Wilson, M. Z. A scalable, easy-to-deploy protocol for cas13-based detection of SARS-COV-2 genetic material. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8092748/. (Accessed Jan 30, 2022).
54.Mustafa, M. I.; Makhawi, A. M. Sherlock and DETECTR: CRISPR-CAS systems as potential rapid diagnostic tools for emerging infectious diseases. https://journals.asm.org/doi/10.1128/JCM.00745-20#B31. (Accessed Jan 30, 2022).
55.Mehta, A.; Merkel, O. M. Immunogenicity of Cas9 protein. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7115921/ (accessed Jan 30, 2022).
56.Abdelhak, S.; Chadwick, R.; Clarke, P.; Gänsbacher, B.; Ghaly, M.; Herbert, M.; Jaenisch, R.; Kato, K.; Lewens, T.; Lim, J.; Liu, D.; Lo, D.; Lovell-Badge, R.; Naldini, L.; Niakan, K.; Ormond, K.; Terry, S. Heritable Human Genome Editing. National Academies Press 2020. (Accessed Jan 30, 2022)

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Development of Hydrogel-based Wound Healing Patch with 6-Gingerol for Animals

Jennifer S. Yu
Yongsan International School of Seoul, 285 Itaewon-ro, Yongsan-gu, Seoul, 04347, Republic of Korea; dgosyber1@gmail.com

ABSTRACT: Agarose is a material extracted from algae and has a self-gelling ability and flexible mechanical properties. Since it is odorless and edible, agarose can be applied to wound healing patches for animals. To develop an agarose-based hydrogel patch, the concentration of agarose supplemented with glycerol, a non-toxic plasticizer was optimized. After testing the mechanical properties such as flexibility, transparency, stiffness, and tensile strength, a patch with 3% agarose and 10% glycerol and a 0.186 cm thickness was found to be the optimal condition for the hydrogel patch. Then, the hydrogel patch was applied to a live dog's leg and was found to be durable for at least 48 hours. Finally, 6-gingerol, a candidate for an anti-inflammatory substance to be embedded in the hydrogel, was tested. As a result, 6-gingerol effectively decreased the mRNA expression level of Cxcl5, a pro-inflammatory marker, in B16F1 mouse skin cells, *in vitro*. In conclusion, an agarose hydrogel patch can be used as a wound dressing sheet for animal skin, and 6-gingerol may be added into the hydrogel patch as an effective anti-inflammatory agent for epidermal wounds.
KEYWORDS: Cellular and molecular biology; Wound healing; Hydrogel; Inflammation; 6-gingerol.

Introduction

Agarose, a material extracted from algae sources, is a water-soluble substance that has a self-gelling ability and flexible mechanical properties. Initially a white powder, agarose must undergo induction, gelation, and quasi-equilibrium to form the commonly known agarose gel. Its permeability and function can be modified by using different concentrations and blending other substances into the gel. Due to its promising characteristics, agarose has been commonly used and is expected to further expand its application in the field of biomedical research.[1]

Hydrogels are cross-linked 3D networks with hydrophilic structures, which allow them to hold large amounts of water. They have many desirable properties such as biocompatibility, biodegradability, and porous structure. Furthermore, different features of hydrogels can react to different stimuli, such as temperature, pH, and light, which can be used to mimic various *in vivo* environments. Hydrogels can be created from natural resources, such as polysaccharides, and synthetic materials. However, some hydrogels can be very brittle. In order to overcome some of the existing limitations, numerous studies are being carried out to develop novel hydrogels with stronger and more stable characteristics while preserving many of the advantages of conventional hydrogels.[2]

Herbal medicine, also known as botanical medicine, is folk medicine made from plants and plant extracts. It has been traditionally used for about 60,000 years. After the discovery of aspirin and the development of chemical medicine, reliance on herbal medicine had greatly decreased. However, as bacterial resistance to antibiotics has risen and the need for more eco-friendly medications increases, many scientists have turned to plant-based materials for solutions. Much research is being done on the superiority and feasibility of herbal medicine compared to chemical medicine. Thus, public acceptance of herbal medicine has significantly increased, and many scientists believe that herbal products will continue to play a crucial role in the health care system.[3] 6-gingerol is a pharmacologically active component abundantly contained in ginger. It has anti-inflammatory and antibacterial properties, making it an optimal compound for wound healing.[4] It is also able to regulate multiple targets. Furthermore, gingerol suppresses carcinogens in the skin of animals. Due to its favorable qualities and safety, 6-gingerol is likely to be researched further as a therapeutic agent.

Methods

Optimization of agarose-glycerol hydrogel:

Different concentrations of agarose ranging from 0.1 to 5% were added to a glass bottle. Then it was heated in a microwave for 80 seconds. 18 ten-centimeter circular plastic plates were labeled by volume and concentration of agarose. Solutions were then heated and poured into plastic dishes and allowed to cool until the gels became fully solidified. After the gelation, the plastic plates were placed in a 50 °C drying oven for 24 hours. The same procedure was performed for 0, 2, 4, 6, 8, 10% of glycerol in a 3% agarose solution.

Fabrication of hydrogel patches using cellulose fibers:

After the hydrogel patches were synthesized, they were placed on the cellulose fiber. The cellulose fibers and hydrogel patches were microwaved for 1 minute to slightly smear the hydrogel patches onto the cellulose fibers. The hydrogel patches were allowed to cool down to room temperature for 10 min.

B16F1 cell culture and LPS treatment:

5 mg of lipopolysaccharide (LPS) was measured using an electronic densimeter and weight boat. 5 mL of Dulbecco's modified phosphate buffer saline (DPBS) was used to dissolve the 5 mg of LPS. Then, the solution was vortexed for 5 minutes to completely dissolve LPS. 500 microliters of trypsin ethylenediaminetetraacetic acid (EDTA) were used to detach B16F1 cells from the culture plate. After centrifugation, cells

DOI: 10.36838/v4i4.12

were resuspended in RPMI-1640 medium. Cells were then treated with different concentrations of LPS and culture for 24 hours in a 37 °C incubator with 5% CO_2.

R condition for Cxcl5:

Polymerase Chain Reaction (PCR) was performed using the Premix PCR kit (Bioneer). 20 μL PCR was prepared to amplify the target cDNA (Cxcl5 and beta-actin). The following PCR condition was used for the amplification of Cxcl15 and beta-actin. Step 1: 95 °C for 3min, step 2: 95 °C for 30 sec, 62 °C for 30 sec, 72 °C for 20 sec, step 2 was repeated for 34 cycles, step 3: 72 °C for 5 min, and step 4: infinite 12 °C.

Results and Discussion

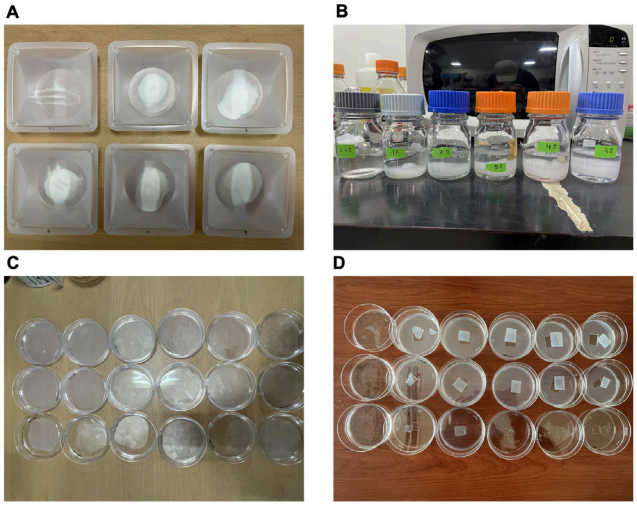


Figure 1: Four steps to prepare agarose hydrogel: A) The weight of the agarose powder was measured by a densitometer. B) Deionized water was added to the agarose powder and microwaved to dissolve completely. C) The dissolved agarose solution was poured into the petri dish. D) The center of the agarose gel was removed to measure tensile strength and plasticity.

The purpose of the first experiment was to find the optimal concentration and thickness of agarose hydrogel for the wound healing patch (Figure 1). 18 samples of agarose hydrogel were prepared to find the optimal conditions for the hydrogel patch. High concentrations and thick agarose hydrogel may increase tensile strength and stiffness. Low concentrations and thin agarose hydrogel may increase the flexibility of the hydrogel.

Table 1: Physical properties of agarose gel with different agarose concentrations.

Agarose (%)	Gel solidification	Tensile strength	Flexibility	Transparency	Stiffness
0.1	No	X	N/A	O	X
1	Yes	X	O	O	X
2	Yes	Δ	O	O	Δ
3	Yes	O	O	O	Δ
4	Yes	O	Δ	Δ	O
5	Yes	O	Δ	Δ	O

To find out the best concentration of agarose gel, six different concentrations of agarose hydrogel were compared. The agarose gel with 0.1% did not solidify. Although the 1% agarose gel solidified, it significantly lacked tensile strength and stiffness, causing it to rip easily. The 2% agarose gel was more durable, but it still lacked tensile strength and stiffness. 4% and 5% were both highly durable but lacked transparency and flexibility. Thus, the 3% agarose gel was found most optimal because it had balanced characteristics between tensile strength, flexibility, and stiffness. (Table 1)

Table 2: Physical properties of agarose gel with different thicknesses.

Thickness of agarose hydrogel	Gel solidification	Tensile strength	Flexibility	Transparency	Stiffness
0.063 cm	O	Δ	O	O	Δ
0.126 cm	O	O	O	O	O
0.189 cm	O	O	Δ	O	O

To find out the best thickness of the hydrogel, 3 different thicknesses were compared. The 0.063 cm agarose gel lacked tensile strength and stiffness, causing it to rip easily. The 0.189 cm agarose gel lacked flexibility making it hard to apply to the wound area. Thus, the 0.126 cm agarose gel was found optimal due to its well-rounded characteristics (Table 2).

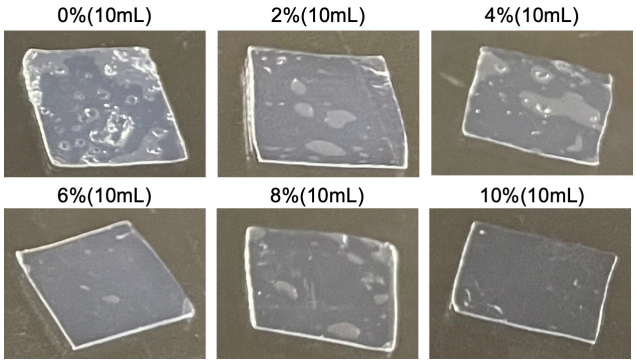


Figure 2: Images of 3% agarose hydrogels with the addition of different concentrations of glycerol.

Table 3: Physical properties of agarose gel with different concentrations of glycerol additive.

Glycerol concentration	Bubble	Flexibility	Surface smoothness	Tensile strength	Adhesiveness
0%	O	X	X	X	X
2%	O	X	X	X	X
4%	O	Δ	X	X	X
6%	Δ	Δ	Δ	X	Δ
8%	Δ	O	Δ	Δ	O
10%	X	O	O	O	O

To find the optimal concentration of glycerol, six different concentrations were tested (Figure 2). Concentrations below 10% formed bubbles and lacked the required physical characteristics. 10% glycerol had no bubbles with the high

est flexibility, surface smoothness, tensile strength, and adhesiveness, making it the optimal concentration (Table 3).

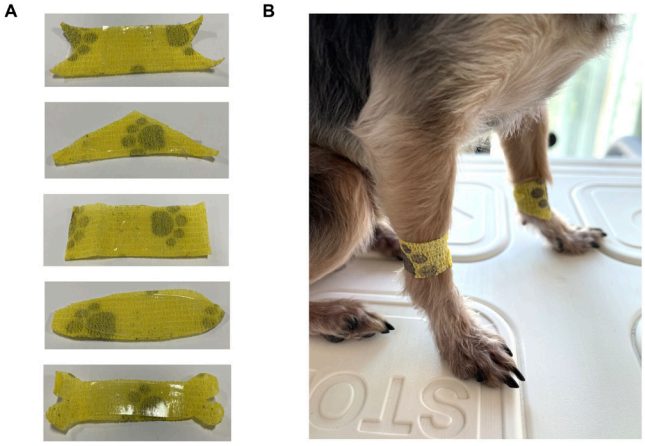


Figure 3: Application of hydrogel using cellulose patches. A) five different conceptual designs of hydrogel patches for the different application areas. B) Application of hydrogel patch on a dog's leg.

After the hydrogel was placed on the cellulose fiber, it was heated by microwave for 1 minute to slightly melt the hydrogel patches on the cellulose fibers. After 10 minutes of cooling at room temperature the hydrogel and cellulose patches were fused together (Figure 3A). These patches were applied to a live dog's leg to test its durability. It was found that the patches remained adhered to for at least 48 hours. (Figure 3B).

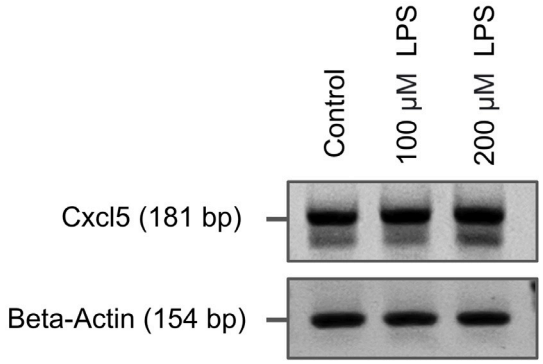


Figure 4: The mRNA expression level of Cxcl5 under different concentrations of LPS treatment.

To mimic an acute inflammation condition, different concentrations of LPS (100 and 200 µM) were treated on B16F1 cells for 24 hours. To confirm the induced inflammatory response, the expression level of C-X-C chemokine (Cxcl5), a neutrophil-activating inflammatory peptide, was measured using PCR. This gene encodes a protein that is a member of the CXC subfamily of chemokines. Cxcl5 chemokines recruit and activate leukocytes. Figure 4 indicates that 200 µM of LPS significantly increased the expression of Cxcl5 in B16F1. Beta-actin was used as the control housekeeping gene.

To test the anti-inflammation effect of 6-gingerol, Cxcl5 expression levels were compared for LPS-induced inflammatory cells. (Figure 5) As expected 20 µM of 6-gingerol decreased the Cxcl5 expression level compared to LPS-induced cells without 6-gingerol treatment. Previous

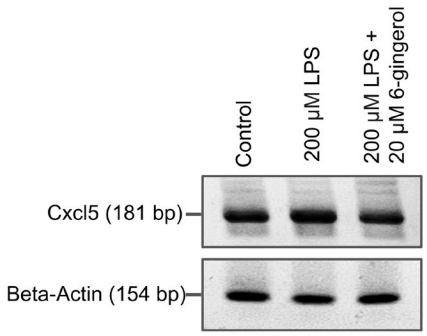


Figure 5: The mRNA expression level of Cxcl5 under LPS and 6-gingerol treatment.

6-gingerol decreased the Cxcl5 expression level compared to LPS-induced cells without 6-gingerol treatment. Previous research indicates that induction of Cxcl5 can be mediated by UVB irradiation in skin. It has been also demonstrated that Cxcl5 causes mechanical pain-related hypersensitivity and induces infiltration of neutrophils and macrophages.[5] Therefore, Cxcl5 may induce the inflammation on skin by UVB irradiated skin cells. Thus 6-gingerol treatment may alleviate inflammation reaction and may reduce pain after skin injury.

Conclusion

An agarose-based hydrogel patch for animal skin wounds was developed in this study. The optimal composition of hydrogel was found to be 3 % agarose with 10 % glycerol. It was also found that 6-gingerol, a natural anti-inflammatory substance, can inhibit Cxcl5-dependent inflammatory pathways. Therefore, the addition of 6-gingerol to agarose hydrogel may reduce inflammation and reduce damage to animal skin tissues. However, more skin cell lines should be tested to verify the anti-inflammatory effect of 6-gingerol. Also, *in vivo* mouse experiments should be performed to check the wound healing effect of this agarose healing patch.

Acknowledgements

I would like to give a special thanks to Dr. Woo Rin Lee for his guidance on this project.

References

1. Varoni, E.; Tschon, M.; Palazzo, B.; Nitti, P.; Martini, L.; Rimon dini, L., Agarose gel as biomaterial or scaffold for implantation surgery: characterization, histological and histomorphometric study on soft tissue response. *Connective tissue research* **2012**, *53* (6), 548-554.
2. Kamoun, E. A.; Kenawy, E.-R. S.; Chen, X., A review on polymeric hydrogel membranes for wound dressing applications: PVA-based hydrogel dressings. *Journal of advanced research* **2017**, *8* (3), 217-233.
3. Dorai, A. A., Wound care with traditional, complementary and al ternative medicine. *Indian Journal of plastic surgery* **2012**, *45* (02), 418-424.
4. Hwang, Y.-H.; Kim, T.; Kim, R.; Ha, H., The natural product 6-gingerol inhibits inflammation-associated osteoclast differen tiation via reduction of prostaglandin E2 levels. *International Jour nal of Molecular Sciences* **2018**, *19* (7), 2068.
5. Dawes, J. M.; Calvo, M.; Perkins, J. R.; Paterson, K. J.; Kiesewetter, H.; Hobbs, C.; Kaan, T. K.; Orengo, C.; Bennett, D. L.; McMahon, S. B., CXCL5 mediates UVB irradiation–induced pain. *Science translational medicine* **2011**, *3* (90), 90ra60-90ra60.

DOI: 10.36838/v4i4.12

Authors

Jennifer Yu is a senior at Yongsan International School of Seoul in South Korea. She actively participates in school activities such as clubs and sports. She is a member of the YISS Varsity Swimming team. She is also the president of Red Cross Youth, the treasurer of Mu Alpha Theta, and a member of the Science Club. She plays the cello and performs in YISS' concerts. In the future, she wishes to major in molecular biology and study at medical school to continue her interests in science and medicine. Additionally, she would like to continue researching various types of topics in college and graduate school.

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

The Cost of Healthy Drinking Water

Fatima A. Khan
Hinsdale Central High School, 5500 S Grant Street, Hinsdale, IL, 60523, United States; fatimakhanhp@gmail.com

ABSTRACT: The purpose of this experiment was to determine what qualities constitute the healthiest drinking water and if the most expensive drinking water was the healthiest for a person to consume. Six types of commonly consumed drinking water products were tested (Smartwater™, Ice Mountain™, Evian™, Dasani™, Fiji™, and DuPage County tap water) for pH, TDS (total dissolved solids, ppm), heavy metals level (ppb), nitrate level (ppm), and chlorine level (ppm). A cost analysis of the different water types was also completed.

The hypothesis of this experiment was not supported completely, as even though the most expensive water, Smartwater™, a distilled water ($8.86 per 100 fl oz), was the highest quality water type, the least expensive water, DuPage County tap water ($0.96 per 100 fl oz), was not the lowest quality water type. Furthermore, the second most expensive water, Evian™, a mineral water ($5.90 per 100 fl oz), is the lowest quality water type. Therefore, it can be seen that the most expensive drinking water is not always the water of the highest quality.

KEYWORDS: Health Sciences; Water; pH; Total Dissolved Solids.

Introduction

There are 3.9 trillion gallons of water consumed per month in the United States.[1] Healthy drinking water is vital to maintaining the human body, such as proper blood circulation and healthy skin. Therefore, it is essential to choose the healthiest drinking water to take care of the body and its systems. Consuming water of poor quality can lead to conditions such as kidney stones and hardened arteries.[2] With a myriad of drinking water for consumers to choose from, it is essential to determine what qualities constitute the healthiest drinking water and whether the most expensive water is the healthiest to consume.

Water quality is defined in terms of the physical, chemical, and biological content of the water. The Environmental Protection Agency (EPA) sets the minimum standards for tap drinking water quality in the United States under the Safe Drinking Water Act of 1974. It is important to note that this has not been updated for twenty-four years, the last amendment being in 1996.[3] Contaminants of water can be physical, chemical, biological, or radiological.[4] Any contaminant, if consumed at a certain level, can be harmful. Tap water comes from surface water, such as streams, rivers, lakes, and reservoirs, or groundwater, which is found beneath the surface. The EPA's standard consists of basic filtration (removing large particles), flocculation (chemicals added to the water to remove smaller particles), and chlorine to kill bacteria and microorganisms.[5] The government does require reducing specific contaminants for safety; however, it is up to the individual state to monitor their drinking water.[6] Most cities do not monitor or filter for medications, therefore these could be present in tap water. Contaminants found in municipal water include prescription drugs, as people dispose of unwanted medicines into the toilet, other drug residues, radioisotopes, fluoride, disinfection chemical additives, and by-products, heavy metals, agricultural contaminants, and sediments.[7] The Food and Drug Administration (FDA) oversees bottled water. People presume that bottled water is healthier, however in 2010, over 173 bottles of different water brands failed the Environmental Working Group's (EWG) transparency tests on contaminants.[8] Contaminants of bottled water include chloride, dissolved solids, and fluoride.[4] Bottled water may be considered less contaminated than tap water, however, no regular testing is required for bottled water manufacturers. It is also important to note that, according to the Beverage Marketing Association, 50% of bottled water is actually bottled tap water.[9] It can, therefore, be seen that the decision to drink a particular type of drinking water is one that needs careful consideration.

According to the CDC, water quality indicators include pH and the level of contaminants.[10] The EPA sets standards and regulations for the presence and levels of over 90 different contaminants.[4] For this study, the most common contaminants will be assessed; chlorine, nitrate, and heavy metals will be examined together with the Total Dissolved Solids (TDS) level and pH level. The pH, or potential hydrogen, is an essential quality of drinking water. Acidic water can cause demineralization of the tooth enamel leading to dental erosion.[11] Conversely, when alkaline water is consumed for a few weeks, it can affect the acid in the human digestive tract leading to malabsorption and growth retardation.[12] The World Health Organization (WHO) warns that extreme pH levels in drinking water can worsen existing skin conditions as well as adversely affect other body systems.[13] The ideal pH of drinking water is 6.5-7.5, which is considered neutral.[14] The addition of chlorine to drinking water is used to kill bacteria and protect consumers from disease, it is also the cheapest way to disinfect water. However, the disinfection process itself can produce harmful by-products. Chlorine in drinking water has been linked to cancers of the bladder, liver, pancreas, and urinary tract.[15] Even though chlorine can eliminate a

DOI: 10.36838/v4i4.13

multitude of unfavorable bacteria, its content in drinking water needs to be carefully monitored, and the EPA recommends that the chlorine level should be less than 1 ppm.[16] Nitrates are a frequent contaminant found in drinking water. Excessive nitrates in drinking water can cause harm to the body, such as deficiencies in fetal growth and child development as well as certain cancers. EPA health guidelines for the level of nitrates in drinking water is less than 1 ppm.[16] Heavy metals present in drinking water can include arsenic, cadmium, nickel, mercury, chromium, zinc, and lead.[17] Although traces of certain heavy metals such as calcium are common and not harmful, heavy metals such as aluminum are linked to diseases. Aluminum in water has been linked to Alzheimer's disease, learning disabilities in children, and gastrointestinal disease.[18] As evident from the Flint water crisis that began in 2016, lead is a heavy metal that can enter the water system through corrosive pipes. It is quickly absorbed, leading to lead poisoning, and children are more vulnerable than adults. Ideally, the heavy metal concentration in drinking water should be less than 10 ppb (micrograms per liter).[16] The lower the TDS level, the purer the water. Increased TDS levels can make water have an unpleasant odor, and make water taste bitter, salty, or metallic. The ideal drinking water should have a TDS of zero to fifty ppm.[16] Using these parameters for the quality of healthy water, an assessment can be made as to which water is the healthiest to consume.

The U.S. National Health Nutrition Examination Survey for 2012, stated the average daily consumption of water for an adult man was 117 ounces (about 15 glasses) and 93 ounces (about 12 glasses) for an adult woman.[19] Popular drinking waters are tap water, Dasani™ (Purified- water that has been mechanically filtered or processed to remove impurities), Fiji™ (Artisan or well water- water that flows under natural pressure without pumping), Smartwater™ (Distilled- water freed of dissolved or suspended solids), Evian™ (Mineral- water from a mineral spring that contains various minerals), and Ice Mountain™ (Spring- water obtained directly from an underground spring).[20] Consumers usually choose the water they drink based on perceived safety and they presume that it is of good quality. With a large choice available to consumers, this study will focus on the most popular brands of bottled water and the tap water of the county the research is being conducted in, DuPage County, Illinois.

The cost of water can vary considerably based on factors such as availability and demand. Consumers also look for health benefits of water and the cost of treatments required to meet those standards has an effect on the price as well. For this study, the average cost of the various brands of water will be taken from the two largest retailers in the United States that carry the same brand and bottle size in DuPage County, Illinois. The tap water price will be taken as an average cost of water throughout DuPage County as published on the DuPage County website.

People will be more conscious of spending their money on the most beneficial water for consumption instead of choosing the most accessible or well-marketed water if they are more aware of the qualities of the water. This study will assess the pH, chlorine level, nitrate level, heavy metal level, and TDS level in six commonly consumed types of drinking water to determine whether the most expensive water is in fact the healthiest to consume.

Methods

Materials needed to test each brand/type of water were; one glass beaker labeled with the brand/type of water, stopwatch, pair of tweezers, five bottles of each water type, one digital TDS meter, one digital pH meter, fifty heavy metal level test strips, fifty nitrate level test strips, fifty chlorine test strips, water (DuPage county tap water, Ice Mountain (spring water), Dasani (purified water), Smartwater (distilled water), Evian (mineral water), Fiji (artesian or well water)). The glass beakers were labelled with the brand/type of water to prevent cross-contamination.

40 mL of the water type being tested was poured into a glass beaker. This was used for each of the five tests (TDS, pH, heavy metals, chlorine, and nitrate) samples. For the TDS test, a digital TDS meter was used. Total Dissolved Solids (TDS) is the combined content of all inorganic and organic substances contained in a liquid. Therefore, anything other than water, such as salt, metals, or minerals, would be included. The TDS meter was immersed into the sample water until the numbers on the screen were steady, and the reading recorded. The probe was cleaned with a paper towel and the prior steps repeated nine more times for each water sample. The glass beaker was then emptied and rinsed with 20 mL of distilled water. 40 mL of the water type being tested was poured into a glass beaker and steps repeated for this sample. The complete procedure for TDS level was repeated for each bottle/sample (a total of 5 bottles/samples per type were tested).

Figure 1: TDS meter

For the pH test a digital pH meter was used and cleaned using distilled water. The digital pH meter was placed into the water sample up to where the end of the cap would sit (the immersion level). It was immersed for one minute (timed using a stopwatch). The reading was recorded when stable. These steps were repeated nine more times for each water sample. The glass beaker was rinsed with 20 mL of distilled water. 40 mL of the water type being tested was poured into a glass beaker and steps repeated for this sample. The complete procedure for pH level was repeated for each bottle/sample (a total of 5 bottles/samples per type were tested)

For the heavy metals test, one heavy metal test strip was dipped into the water sample for thirty seconds with a

Figure 2: pH meter

constant, gentle back and forth motion. The strip was removed and left for 2 minutes before matching the color of the strip to the color chart (located on the heavy metal strips bottle). The color was matched within 30 seconds. The steps were repeated nine more times. The glass beaker was emptied and rinsed with 20 mL of distilled water. 40 mL of the water type being tested was poured into a glass beaker and steps repeated for this sample. The complete procedure for heavy metals level for each bottle/sample was repeated (a total of 5 bottles/samples per type were tested).

Figure 3: Heavy metal testing and color matching

For the nitrate test, one nitrate strip from the nitrate testing strips container was dipped into the water sample for one second. The strip was removed, and color assessed with the pad side up after 30 seconds (the color chart was located on the chlorine test strip bottle). Steps were repeated nine more times. The glass beaker was rinsed with 20 mL of distilled water. 40 mL of the water type being tested was poured into a glass beaker and repeated for this sample. The complete procedure for nitrate level for each bottle/sample was repeated (a total of 5 bottles/samples per type were tested).

Figure 4: Nitrate testing and color matching

For the chlorine test, a chlorine test strip was dipped into the water sample and moved back and forth for 30 seconds. The strip was removed from the water and compared with the color charts on the bottle. These steps were repeated nine more times per sample. The glass beaker was emptied and rinsed with 20 mL of distilled water. 40 mL of the water type being tested was poured into a glass beaker and steps repeated for this sample. The complete procedure for chlorine level for each bottle/sample was completed (a total of 5 bottles/samples per type were tested).

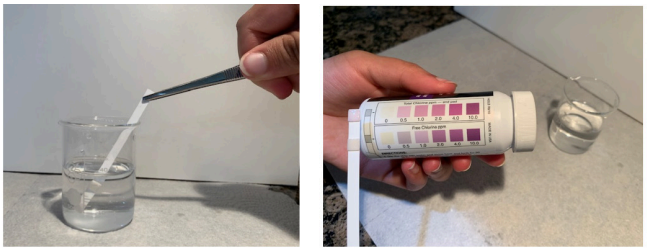


Figure 5: Chlorine testing and color matching

QUALITY ASSESSMENT SCORECARD: Each water type was given a score out of five. One check was given for each parameter that fell within the EPA guidelines. If the parameter being tested did not fall within the EPA guidelines, a score of zero was given for that category.

Results and Discussion

Cost Analysis::

The mean cost per 100 fl oz of different types of commonly consumed drinking water was calculated using prices of the two largest retailers in DuPage County: Target and Walmart (see Table 1).

Table 1: Mean cost of each type of water type in dollars per 100 fl oz.

Type of Water	***Mean Cost ($) per 100 fl oz.***
Smartwater™	$8.86
Ice Mountain™	$1.77
Evian™	$5.90
Dasani™	$1.11
Fiji™	$5.57
DuPage County tap water	$0.96

Discussion of Cost Data:

To reduce the error in assessing the cost of different types of water, the average price of a 16.9 fl oz water bottle for each type of water in two different major retailers in DuPage County, IL (Target and Walmart) was calculated. For DuPage County water, information from the county website was used. The cost of bottled water varies considerably, even within the same water type. For example, the same size bottle of Dasani water may be different if purchased in Walmart vs. Target. The price of drinking water can also be dependent on the cost of production, as more regulations are met for safety, as well as the demand for the particular type of drinking water. During Hurricane Irma in Florida, bottled water was being sold for $99.99 per case (U.S.A. Today, 9/9/17). It was, therefore, evident that there would be some variation in cost, but it was quite surprising to find that there was a difference/range in the

DOI: 10.36838/v4i4.13

cost of drinking water of $7.90 per 100 fl oz between the most expensive (Smartwater™) and the cheapest water (DuPage County tap water). This, therefore, shows that there is indeed a cost factor when choosing which drinking water to consume and further validates the need for a study such as this so that the consumer can make an informed decision about which water to consume. Table 2 below shows the rank order of the type of water from most expensive to least expensive.

Table 2: Table ranking the type of water type from most expensive to least expensive in dollars per 100 fl oz.

Type of Water	*Mean Cost ($) per 100 fl oz.*
Smartwater™	$8.86
Ice Mountain™	$1.77
Evian™	$5.90
Dasani™	$1.11
Fiji™	$5.57
DuPage County tap water	$0.96

pH Data:

The pH of different types of commonly consumed drinking water was obtained. Each type of water was tested 50 times and the mean and standard deviation were calculated for each type of water (see Table 3; Figure 1).

Table 3: Mean pH for the different types of drinking water.

	Smartwater	Ice Mountain	Evian	Dasani	Fiji	DuPage County Tap Water
Mean pH	6.86	7.01	7.61	6.18	7.56	7.74
Standard deviation	0.081	0.033	0.039	0.066	0.049	0.067

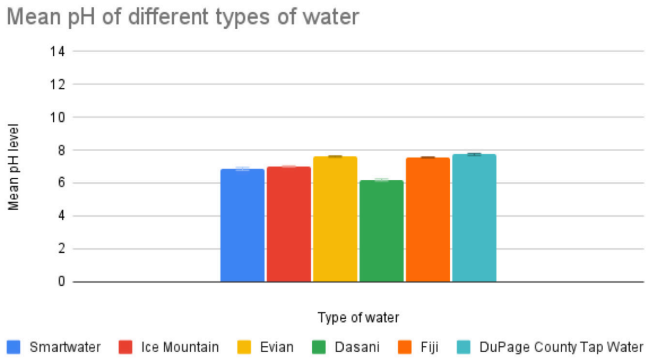


Figure 1: Bar chart displaying the mean pH for each type of drinking water.

It can be seen that the different water types have varying pH levels. Dasani™, a purified water, was the most acidic, having a mean pH of 6.18. It also fell below the EPA guidelines for optimal pH which is a pH of 6.5-7.5. DuPage County water was the most alkaline at a pH of 7.74 and was also not within the optimal range for drinking water. An ANOVA test was conducted to determine if there was a significant difference between the mean pHs of each water type so that further analysis could be conducted about the individual pHs.

pH Discussion:

After completing the ANOVA test, it could be determined that the pH of each individual type of water is statistically significant when compared to the others, p-value <0.05. This means that each type of water has its own individual pH level and there is no overlap with any of the other water types. Dasani™, a purified water, was the most acidic and has an average pH of 6.18. Two brands were in the EPA optimal pH range for drinking water. They were Smartwater™, a distilled water at a mean pH of 6.86, and Ice Mountain™, a spring water with a mean pH of 7.01. DuPage County tap water was the most alkaline with an average pH of 7.74. This together with Fiji™, an artesian water with a mean pH of 7.56, and Evian™, a mineral water at a pH of 7.61, also are considered more alkaline and are above the EPA guidelines for pH (pH 6.5-7.5). The results for pH show that there is a wide range of pH values and that different drinking water types have distinct pH levels.

TDS Data:

This experiment was conducted and the TDS of different types of commonly consumed drinking water was obtained. Each type of water was tested 50 times and the mean and standard deviation were calculated for each type of water (see Table 4; Figure 2).

Table 4: Mean TDS (ppm) for the different types of drinking water.

	Smartwater	Ice Mountain	Evian	Dasani	Fiji	Dupage County Tap Water
Mean (ppm)	20.98	108.24	235.42	23.6	121.32	112.52
Standard deviation	0.74	6.34	16.48	0.73	2.73	5.02

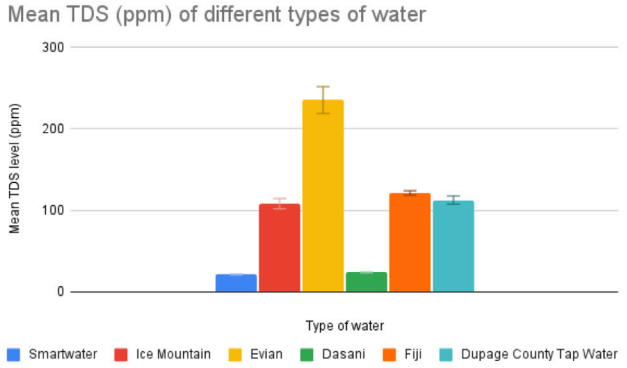


Figure 2: Bar chart displaying the mean TDS (ppm) for each type of drinking water.

It can be seen that the different water types have varying TDS levels. Evian™, a mineral water, had the highest mean TDS level of 235.42 ppm. This is above the EPA guidelines for the optimal TDS level, which should be less than 50 ppm. Smartwater™, a distilled water, had the lowest mean TDS level of 20.98 ppm and was within the EPA recommended guidelines. An ANOVA test was conducted to determine if there was a significant difference between the mean TDS levels of each water type so that further analysis could be conducted about the individual TDS values of each water type.

TDS Discussion:

A single factor ANOVA test was conducted, and it can be concluded that the TDS of each individual type of water is statistically significant when compared to the others, p-value <0.05. This means that each type of water has its own individual TDS level and there is no overlap with any of the other water types. TDS stands for Total Dissolved Solids, which accounts for anything in the water besides water molecules themselves. It is not surprising that the different water types had different

TDS levels (for example, people would expect mineral water to have a high TDS because of minerals in it and they would expect a low TDS from distilled water). However, four of the waters that were tested fell outside of the EPA's recommended TDS level (<50ppm) and were over double the recommended guideline level (DuPage County water- 112.52 ppm, Fiji™- 121.32 ppm, Ice Mountain™- 108.24 ppm) and, in one case, over four times the recommended value (Evian™- 235.42 ppm). It is not surprising that Evian™ had the highest TDS value, as Evian is a mineral water, and these minerals added to the TDS level, as they are not water molecules. To be able to determine a more detailed conclusion for TDS, it is necessary to conduct further testing to establish the exact composition of the water that accounts for the high TDS levels.

Heavy Metal Data:

This experiment was conducted and the heavy metal level of different types of commonly consumed drinking water was analyzed. Each type of water was tested 50 times and the mean and standard deviation were calculated for each type of water (see Table 5; Figure 3).

Table 5: Average heavy metal level (ppb) for the different types of drinking water.

	Smartwater	Ice Mountain	Evian	Dasani	Fiji	Dupage County Tap Water
Mean	10	12	44	10	20	39.2
Standard deviation	0	4.04	12.12	0	0	14.55

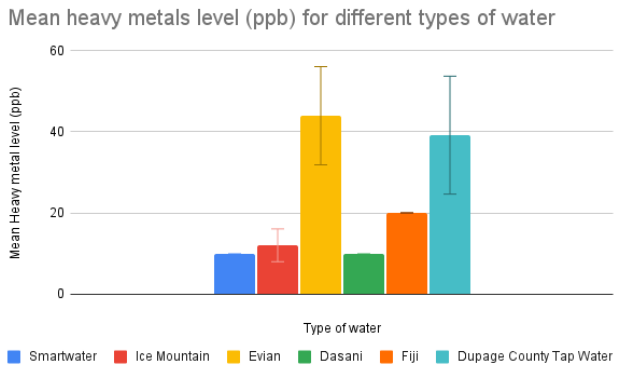


Figure 3: Bar chart displaying the mean heavy metal level (ppb) for each type of drinking water.

Heavy Metal Data Analysis and Discussion:

Four out of the six water types tested had greater than the EPA recommended guidelines level for heavy metals (less than 10 ppb). In descending order, they are Evian™, a mineral water (44 ppb), DuPage County tap water (39.2 ppb), Fiji™, an artesian or well water (20 ppb), and Ice Mountain™, a spring water (12 ppb). The two water types that fell within the EPA guidelines were Dasani™, a purified water (less than 10 ppb) and Smartwater™, a distilled water (less than 10 ppb). The heavy metal test that was used detects the heavy metals Copper, Cobalt, Zinc, Cadmium, Nickel, and Mercury. Therefore, it may be possible that the actual heavy metal content may be higher, as there may be other heavy metals that are in the water that are not being tested for in this test. It is interesting to see that Evian™, a mineral water, actually contained more heavy metals than DuPage County tap water, as Evian™ water is tapped at the source, whereas Dupage County tap water has to run through pipes in order to get to its location. Usually, heavy metals enter the water system through pipes that have some degree of leaching due to corrosion, which may account for the high levels seen in DuPage County tap water (39.2 ppb).

Nitrate Data:

This experiment was conducted and the nitrate level of different types of commonly consumed drinking water was obtained. Each type of water was tested 50 times and the mean and standard deviation were calculated for each type of water (see Table 6; Figure 4).

Table 6: Mean nitrate level (ppm) for the different types of drinking water.

	Smartwater	Ice Mountain	Evian	Dasani	Fiji	Dupage County Tap Water
Mean (ppm)	0	0	1	0	0.5	0
Standard deviation	0	0	0	0	0	0

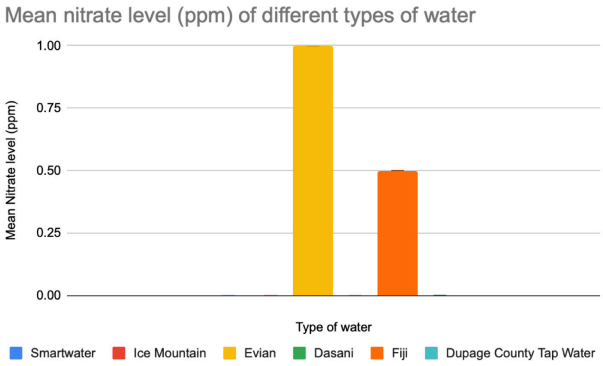


Figure 4: Bar chart displaying the mean nitrate (ppm) for each type of drinking water.

Nitrate Data Analysis and Discussion:

All water types except for Evian™, a mineral water, and Fiji™, and artesian water, had a level of 0 ppm for nitrates. As nitrates are a common contaminant of underground water, it is not surprising that Evian™ water contains some levels of nitrate (1 ppm). Rain and snow flow from the Vizier Plateau to an underground spring, where it is trapped in sandy beds between two layers of clay before being harvested. This may account for the nitrate level seen. Fiji™ water is essentially a spring water sourced from a well dug in the ground. Internal pressure from the ground forces the water into the dug hole and it is collected in an aquifer. Therefore, the water that is collected is underground. Fiji ™water specifically comes from the island of Viti Levu in Fiji and is naturally filtered through volcanic rock. It is, therefore, not surprising that it would contain some nitrates (0.5 ppm). The EPA guidelines state that nitrate levels in drinking water should be less that 1 ppm. Fiji™, an artesian or well water, does fit within the EPA guidelines (0.5 ppm), however Evian™, a mineral water, does not at 1ppm.

Chlorine Data:

This experiment was conducted and the chlorine level of different types of commonly consumed drinking water was obtained. Each type of water was tested 50 times and the mean and standard deviation were calculated for each type of water (see Table 7; Figure 5).

DOI: 10.36838/v4i4.13

Table 7: Mean chlorine level (ppm) for the different types of drinking water.

	Smartwater	Ice Mountain	Evian	Dasani	Fiji	Dupage County Tap Water
Mean (ppm)	0	0	1	0	0.5	0
Standard deviation	0	0	0	0	0	0

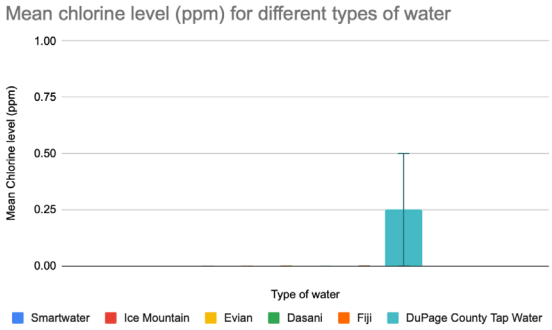


Figure 5: Bar chart displaying the mean chlorine level (ppm) for each type of drinking water.

Chlorine Data Analysis and Discussion:

It can be seen from Figure 5 that out of the waters tested, the only water that contained levels of chlorine was DuPage County tap water. The addition of chlorine to drinking water has been standard since 1904. It is used to kill bacteria and protect users from disease because it is also the cheapest way to disinfect water. However, the disinfection process itself can produce disinfection by-products which can be harmful such as trihalomethanes which are carcinogenic by-products. The DuPage County tap water did meet EPA requirements of having a level of less than 1 ppm of chlorine and monitoring should continue to ensure that the level of chlorine remains within the appropriate guidelines. It is good to know that all the other water types have no traces of chlorine.

Table 8: Table showing the scorecard for each type of water water.

	EPA Guidelines	Smartwater	Ice Mountain	Evian	Dasani	Fiji	DuPage County tap water
pH	6.5-7.5	✅	✅	❌	❌	❌	❌
Heavy metals	<10 ppb	✅	❌	❌	✅	❌	❌
TDS	<50 ppm	✅	❌	❌	✅	❌	❌
Chlorine	<1 ppm	✅	✅	✅	✅	✅	✅
Nitrates	<1ppm	✅	✅	❌	✅	✅	✅
	Score:	5/5	3/5	1/5	4/5	2/5	2/5

Quality Assessment Scorecard:

It can be seen that Fiji™, an artesian or well water, and DuPage County tap water have the same score. To further analyze the difference between the quality of these two waters, the actual parameter values were assessed. Table 9 below shows the mean data for each parameter tested for and highlights those that do not fall within the EPA guidelines.

A comparison of the parameters that did not fit the EPA guidelines for Fiji™, an artesian or well water, and DuPage County tap water was conducted to analyze which water type fell closer to the EPA guidelines. Fiji™ water has a pH of 7.56, which is closer to the EPA guidelines than DuPage County tap water, which has a pH of 7.74. Fiji™ water also has a lower heavy metal content of 20 ppb when compared to the heavy metal content of DuPage County tap water of 39.2 ppb. However, the TDS level for DuPage County tap water (112.52 ppm) is lower than the TDS level of Fiji™ water (121.32 ppm), making this quality more superior in DuPage County tap water. Overall, Fiji™ water has a slightly better quality than DuPage County tap water. Table 10 shows a side-by-side comparison of water quality vs water cost (per 100 fl oz).

Table 9: Table showing the mean value for each parameter tested for in comparison to the EPA guidelines. Those parameters that do not fall within the guidelines are highlighted.

	EPA Guidelines	Smartwater	Ice Mountain	Evian	Dasani	Fiji	DuPage County tap water
pH	6.5-7.5	6.86	7.01	7.61	6.18	7.56	7.74
Heavy metals	<10 ppb	<10	12	44	<10	20	39.2
TDS	<50 ppm	20.98	108.24	235.42	23.6	121.32	112.52
Chlorine	<1 ppm	0	0	0	0	0	0.25
Nitrates	<1ppm	0	0	1	0	0.5	0
	Score:	5/5	3/5	1/5	4/5	2/5	2/5

Comparison of Quality of Water vs. Cost of Water:

A side-by-side comparison of the rank list of quality of water and cost of water is seen in Table 10 below.

Table 10: Table showing a side-by-side comparison of water quality vs water cost (per 100 fl oz).

Rank of Quality	*Type of Water*	*Rank of Expense*	*Type of Water*
1	Smartwater™ (distilled water)	1	Smartwater™ (distilled water)
2	Dasani™ (purified water)	2	Evian™ (mineral water)
3	Ice Mountain™ (spring water)	3	Fiji™ (artesian or well water)
4	Fiji™(artesian or well water)	4	Ice Mountain™ (spring water)
5	DuPage County tap water	5	Dasani™ (purified water)
6	Evian™(mineral water)	6	DuPage County tap water

It can be seen that the most expensive water, Smartwater™, is the water of the highest quality within the types of water tested. However, the cheapest water (DuPage County tap water) was not the worst quality water. Evian™, a mineral water, the second most expensive water type, scored as the lowest quality water due to the fact that it had a higher heavy metal content, TDS level, and nitrate level. Dasani™, a purified water, was the second highest quality water and was the second cheapest water type. It can therefore be seen that there is no direct correlation between water quality and cost (Table 8).

Error Analysis:

Variables in the experiment were controlled by conducting the experiment in the same location and environment. Procedures were followed exactly for 50 trials per water type to obtain a mean value. Each type of water had a specifically labeled glass beaker that was assigned to prevent cross-contamination with other water types. Tools used between water types and samples were thoroughly cleaned using the procedures stated to prevent cross-contamination. The experiment could have been improved by using various different water samples from different locations in DuPage County.

■ Conclusion

The purpose of this experiment was to determine what qualities constitute the healthiest drinking water and if the most

expensive drinking water is the healthiest for a person to consume. For this experiment, the quality of the water was determined by the EPA guidelines for five different qualities of drinking water (pH, TDS, heavy metals, nitrates, and chlorine). In addition to these observations, to address the second part of the purpose, the average cost of each type of water was calculated and compared to the water's quality.

The hypothesis of this experiment stated that if different types of drinking water were tested, the most expensive water would be healthier than its less expensive competitors because it was assumed to have a higher quality. This hypothesis was not supported completely, as even though the most expensive water, Smartwater, a distilled water priced at $8.86 per 100 fl oz, was of the highest quality water type, the least expensive water, DuPage County tap water priced at $0.96 per 100 fl oz, was not of the lowest quality water type. Furthermore, Evian, a mineral water and the second most expensive water type priced at $5.90 per 100 fl oz, was the second-lowest quality water type. Dasani™, a purified water, was the second cheapest water type priced at $1.11 per 100 fl oz and was the second-highest quality water. Therefore, from this study, it can be seen that the highest quality water may not always be the one that is the most expensive.

In conclusion, from the study, it is evident that there are a variety of water types for consumers to choose from and that each water type can vary considerably in its quality regardless of its cost. The highest quality water may not always be the one that is the most expensive and that the lowest quality water may not be the cheapest. Furthermore, due to the significant differences in drinking water quality, it is important for more information detailing the contents of the various types of drinking water to be accessible so that the consumer is aware of what they are drinking. If consumers are more aware of the qualities of healthy drinking water, they will be more conscious of spending their money on the most beneficial water for consumption instead of choosing the most accessible or well-marketed water. It is important to choose the healthiest drinking water to consume in order to take care of the body and its systems.

Acknowledgements

I would like to thank my teachers Mr. Jazak and Mr. Wollschlaeger, for mentoring me throughout the scientific research process and providing their knowledge and advice.

References

1. USGS. (2019, September 23). How Much Water Does the Average Person Use? The Last Well. Retrieved July 8, 2020, from https://thelastwell.org/how-much-water-does-the-average-person-use/
2. Sengupta P. Potential Health Impacts of Hard Water Int J Prev Med 2013;4:866-75.
3. Agency, Environmental Protection. "EPA History: Safe Drinking Water Act." EPA, Environmental Protection Agency, https://www.epa.gov/history/epa-history-safe-drinking-water-act.
4. Agency, Environmental Protection. "Types of Drinking Water Contaminants." EPA, Environmental Protection Agency, https://www.epa.gov/ccl/types-drinking-water-contaminants.
5. "Water Treatment." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 20 Jan. 2015, https://www.cdc.gov/healthywater/drinking/public/water_treatment.html.
6. Centers for Disease Control and Prevention. (2020, November 9). Drinking Water Frequently Asked Questions (FAQs). Centers for Disease Control and Prevention. https://www.cdc.gov/healthywater/drinking/drinking-water-faq.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fhealthywater%2Fdrinking%2Fpublic%2Fdrinking-water-faq.html.
7. "Pharmaceuticals in Water." Pharmaceuticals in Water, https://www.usgs.gov/special-topic/water-science-school/science/pharmaceuticals-water?qt-science_center_objects=0#qt-science_center_objects.
8. Leiba, N., Gray, S., & Houlihan, J. (2012). (rep.). 2011 Bot tled Water Scorecard. Environmental Working Group. Retrieved fromhttps://static.ewg.org/reports/2010/bot tledwater2010/pdf/2011-bottledwater-scorecard-report.pdf?_ga=2.4987120.2137939319.1612822087-1199741675.1612822087
9. Sadler, A. T. (2017, March 22). Half of all bottled water is just tap water. Clark. Retrieved July 8, 2020, from https://clark.com/deals-money-saving-advice/pepsi-aquafina-tap-water-best-bot tled-waters/
10. Centers for Disease Control and Prevention. (2020, November 9). Drinking Water Frequently Asked Questions (FAQs). Centers for Disease Control and Prevention. https://www.cdc.gov/healthywater/drinking/drinking-water-faq.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fhealthywater%2Fdrinking%2Fpublic%2Fdrinking-water-faq.html.
11. Ehlen, Leslie A, TA. Marshall, F. Qian, JS Wefel, & JJ Warren "Acidic beverages increase the risk of in vitro tooth erosion." Nu trition research (New York, N.Y.) vol. 28,5 (2008): 299-303. doi:10.1016/j.nutres.2008.03.0017
12. Merne, M E, KJ Syrjaenen, & SM Syrjaene "Systemic and local effects of long-term exposure to alkaline drinking water in rats." International journal of experimental pathology vol. 82,4 (2001): 213-9. doi:10.1111/j.1365-2613.2001.iep0082-0213-x
13. 8 Chemical Aspects - World Health Organization. https://www.who.int/water_sanitation_health/dwq/gdwq0506_8.pdf.
14. Mercola, J., Dr. (2020, January). How Drinking More Spring or Filtered Water Can Improve Every Facet of Your Health. Mercola Take Control of Your Health. https://www.mercola.com/article/water.htm
15. Hrudey, S., Backer, L., & Humpage, A. (2015). Evaluating Evidence for Association of Human Bladder Cancer with Drink ing-Water Chlorination Disinfection By-Products. Journal of Tox icology and Environmental Health, 18(5), 213–241. https://doi.org/10.1080/10937404.2015.1067661
16. Agency, Environmental Protection. "National Primary Drinking Water Regulations." EPA, Environmental Protection Agency, https://www.epa.gov/ground-water-and-drinking-water/nation al-primary-drinking-water-regulations.
17. Rehman, Kanwal, Fiza Fatima, Iqra Waheed, & Muhammad Sajid Hamid Akash "Prevalence of exposure of heavy metals and their impact on health consequences." Journal of cellular biochemistry vol. 119,1 (2018): 157-184. doi:10.1002/jcb.26234
18. Sullivan, M., & Leavey, S. (2011). Heavy Metals in Bottled Nat ural Spring Water. Journal of Environmental Health,73(10), 8-13. Retrieved February 8, 2021, from http://www.jstor.org/stable/26329231
19. "Products - Data Briefs - Number 242 - April 2016." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 26 Apr. 2016, https://www.cdc.gov/nchs/products/databriefs/db242.htm.

Author

Fatima Khan has been awarded Gold at the science fair regional level three times and has been the recipient of the Illinois

DOI: 10.36838/v4i4.13

State top 10 paper award as well as the Jay & Jerry Hoffman award for excellence in science. In 2021 she was awarded Gold at the state level.

IJHSR
ijhighschoolresearch.org
■ RESEARCH ARTICLE

Effect of Shifting to Virtual Platforms on the Mental Health in School Children

Laisha Gulia
Good Shepherd International School, Ooty, Tamil Nadu, 643004, India; sgulia2000@gmail.com

ABSTRACT: Due to the precautions exercised to control the corona virus disease (COVID-19) pandemic, various population groups have acclimated to diverse levels of screen usage which may have substantial effects on physical and mental health. Numerous studies have been done to identify the screen time across various population segments. Digital screen use which includes television, computers, mobile phones and smart devices has been linked to a variety of health problems. Some of these health problems include obesity, high blood pressure, diabetes, myopia, depression, sleep difficulties and a variety of other non-communicable diseases. This increased disease implication is prominent among those who have an inactive lifestyle which is likely to worsen during quarantine or isolation owing to COVID-19. As a result, it is vital to analyze the long-term health repercussions of such behavior. A study was conducted on various school children for their screen time usage via social media and Google forms. The study leads to the conclusion that the virtual mode of learning has caused various health issues in school children and traditional learning environment should be used to fix the health issues caused by the pandemic.

KEYWORDS: Behavioral and Social Sciences; Psychology; Screen time; Covid-19; Mental Stress; Smart Devices; Health issues.

■ Introduction

Globally, the pandemic has had a devastating effect on physical and mental health of people.[1,2] Governments took a variety of measures to avoid disease transmission which included implementing shelter-in-place laws, encouraging residents to stay at home, restricting entry to nursing facilities and prohibiting gatherings in areas where individuals may come into closer contact.[3] Populations with higher preventative measures, increased contact tracing, and efficient clinical care have had better public health outcomes during this pandemic.[4-6] However, the closure of schools, offices, and other organizations and the shift to online modes has resulted in an increased reliance on digital media such as desktop computers, laptops, and mobile devices for the only means of communication and other activities.[7,8] Individuals who stay at home and are not doing any physical activity spend more time watching television or using digital media for entertainment. Screen usage has been linked to a wide variety of disorders such as anxiety, sleep deprivation, obesity etc. in susceptible populations.[9] Recent empirical investigations indicate that digital screen time has increased rapidly in several communities during the COVID-19 pandemic. This requires a thorough awareness of the potential public health consequences and the preventive policies that should be implemented to mitigate them.[10]

This paper presents a survey to measure the of effects of the pandemic on the screen usage of various people; the main areas of concern were increase in the use of social media, change in the sleep patterns, reasons for the increase of screen time, and adaptability of students to virtual platforms for online classes. Section II describes the methodology used to collect the data. The survey results are discussed in Section III followed by discussion of previous studies and some suggested preventive measures. The last section describes the conclusions.

■ Methodology

The survey was carried out using Google Forms. For maximum participation, the link was shared via different mediums like WhatsApp, Facebook and emails. The participants were mainly a group of students from 15-21 age groups (currently studying) and a sample size of 120 students was taken. The questionnaire in Google Forms consisted of various questions related to their significant shift in screen time, effect of quarantine on screen time, effect of the pandemic on usage of social media and how the increased screen time has affected the sleep patterns and consequently the health of students. The questions in the form were designed based on their relevance in daily lifestyle of children. This made it easier for them to understand and answer. The questionnaire was circulated online, after the first wave of COVID was over and the students had attended one term of online classes. At this stage, the children had already transitioned to online platforms, and they were spending much of their time on screens. Depending on their responses, the inferences were drawn and analyzed. The graphical depiction has been done using Pie Charts.

■ Results

The responses of students are as follows:

a) Effect of Quarantine on screen time: All the students were left to themselves with minimum interaction with their friends, relatives, teachers and even families. The pandemic led to long lasting effects on them such as anxiety, depression, obesity, sleep deprivation, and even lower test grades. Figure 1 shows the effect of pandemic on screen time. It depicts that

DOI: 10.36838/v4i4.14

88.6% of the students who participated in the survey have increased screen time during lockdown. However, 11.4% students feel that their screen time might have increased but by not a significant amount. These students would not have felt a significant shift in their screen time as they might already be involved in some online activities even before the pandemic started.

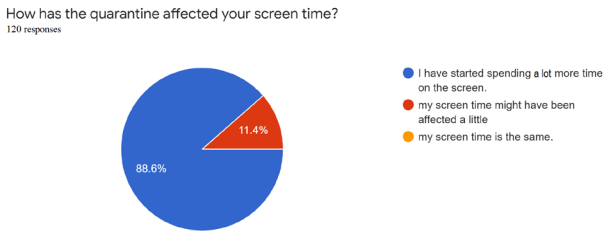


Figure 1: Effect of quarantine on the screen time.

b) The next question asked was about how the current situation had affected their use of social media. Figure 2 shows a pie chart which clearly shows that there is an increase in the time that is spent on the social media. Nearly half of the students felt that they have started using the social media sites more often and they cannot stop using the social media. A total of 22.9% people felt that they spend more than the usual time on social media, but they are trying to control it. Only 34.3% felt that they used to visit such sites for the same amount of time as before. This shows that the students started depending upon the virtual world to keep away their loneliness.

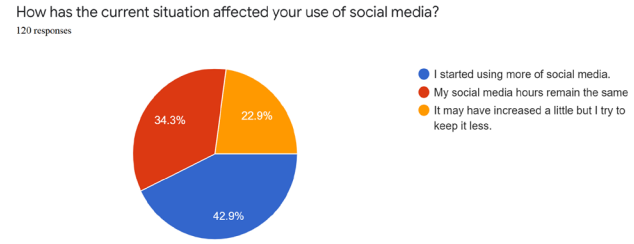


Figure 2: Effect on use of social media due to quarantine.

c) Screen time has various effects on human body. Figure 3 shows effects of screen time on the sleep pattern of the participants. Approximately 63% participants found that their sleep cycle had been disturbed due to the increase in their screen time over the pandemic lockdown, while 37.1% thought that they were able to sleep properly. Although the increased screen time affected the sleep schedule of the students, many students have already had changes in their sleep schedule before everything shifted to a virtual platform and are adapted to it or the change in their screen time did not affect their sleep pattern. When children spend large amounts

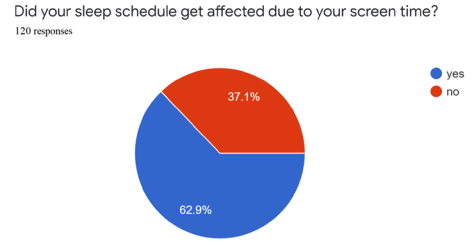


Figure 3: Effect on sleep due to quarantine

of time on screen right before bed, it has been shown to interfere with the sleep quality and the overall sleeping time.[11] This sleep deprivation leads to stress which causes a poor performance in school as well as lower grades.

d) Different people have different effects on their health due to quarantine; approximately 43% people experienced more frequent headaches while 40% of people did not feel any sort of difference between before and after lockdown as shown in Figure 4. As students spend more time on screen, it affects their eyesight. This causes weakness in the eyes. Spending long hours in front of screen leads to disruptive cognitive function and permanent structural changes in the brain.[12] Headaches are also a cause of weak eyesight and increases the risk of myopia.[13]

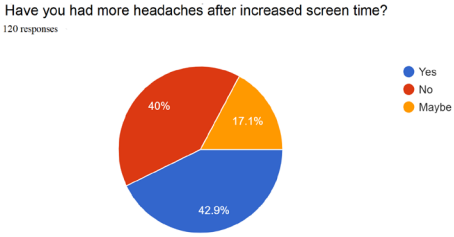


Figure 4: Headache comparison before and after lockdown

e) From Figure 1, it was seen that there was an increase in screen time during the lockdown period for most of the students, so it was essential to know where they were actually spending most of their screen time. Was it because of school/college being conducted online or some other reason? From Figure 5 it can be shown that majority of the students, 77%, believed online classes were the reason for increased screen time. Approximately 14% felt that increased screen time could not be attributed to online classes. Others were also related to studies or work. The online classes and other internships are an important part of students' lives, and these cannot be avoided. Therefore, the students had no choice than to spend most of their time on a screen. This led to many health problems.

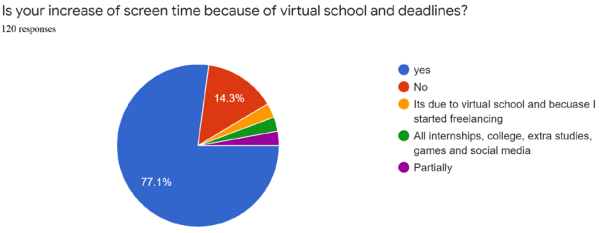


Figure 5: Effect of online classes on screen time

f) Since screen time has increased drastically on average, people have been doing different activities to reduce their screen time. Most people like to utilize their time by doing physical activities such as exercise or dance. Figure 6 shows that 51.4% of the students were not doing any physical activity during the lockdown period and they did not like to spend their time outside the working environment. Only 28.6% of the students thought of engaging themselves in one or the other physical activity. Physical workout is considered the most effective way to get away from the virtual world and to keep the body and the mind healthy. Less amount of time de

voted to physical exercises can lead to future problems like depression and other mental health problems.

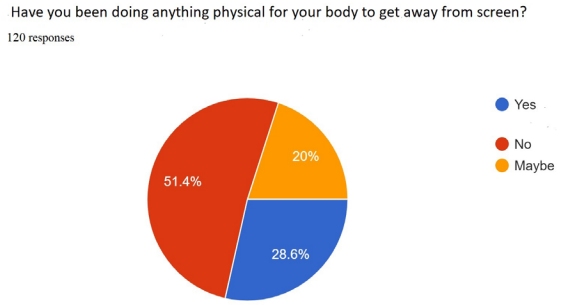


Figure 6: Physical workout during the pandemic

Discussion

A study was conducted on the college undergraduate students about the effects of mobile devices on students revealed that they experienced the symptoms of headaches, anxiety, and irritability by the continuous use of cell phones.[14] The overuse of mobile phones has also led to hearing impairment and a pain in the thumb. This study was conducted long before the end of the pandemic in the world. Another study was conducted to study the relationship between depression and the increased screen time.[15] It concluded that a sedentary lifestyle due to increased screen time leads to higher symptoms of depression and anxiety. The study conducted suggests that with each hour increase in the use of social media apps, the symptoms of depression are also increased.[16] The study was conducted before the pandemic, but it helps in analyzing the mental health of school children whose screen time was drastically increased during the lockdown period. In another study, the effect of lockdowns was studied in young adults.[17] The study shows that more than 50% of the adults reported severe symptoms of depression. It also showed that it affected their sleep patterns, and the wake-up timings was also altered drastically. This study was conducted on the adults while this study focused on the school children who were forced to increase their screen time due to online classes. There was a significant increase in the negative emotions during the pandemic like anxiety, fear and annoyance.[18]

A summary of the analysis drawn from the responses collected is shown in Table I.

Table 1: Inferences Drawn from the Responses.

Question	Response			Inferences
	Yes	No	Maybe	
Are you spending more time on screens?	88.6%	11.4%		Sudden shift to online platform and the inability to interact with their fellow students in person, resulted in increased screen time.
Has your use of social media increased?	43%	34%	23%	Lockdowns forced the children to find friends online.
Has your sleep schedule been affected due to increases in screen time?	63%	37%		Children became addicted to mobile phones and social media. It affected their sleep behavior.
Have you had more headaches than usual after increased screen time?	43%	40%	17%	Overuse of devices disrupted the physical health of children. Poor eyesight and mental stress can be reasons for headaches.
Is your increased screen time due to virtual classes and deadlines?	77%	14%	9%	Sudden changes in the study methods increased the use of devices. Moreover, ease of access to the devices also led to increase in screen time.
Have you been doing any physical workout during the pandemic to get away from the screen?	29%	51%	20%	Less physical workout leads to more physical and mental problems.

of participants' increased screen time was all academic while 38.8% felt that they enjoyed surfing social media during their leisure time. Students needed some recreation and relaxation time which they felt can be attained by spending time on social media.

Children can reduce their screen time by keeping track of screen time and set time limits, by keeping the phone out of the bedroom, by removing unnecessary apps from mobile devices, by switching to grayscale to avoid strain on eyes, and by indulging in a new hobby.

Conclusions

Technological advancements have enabled people to become far more productive. The addition to these new features to electronic devices is very tempting and makes lives easier. Yet it is important to keep the use of devices such as laptops, phones, and televisions under control. The pie charts clearly show that students have had many side effects on their mental as well as their physical health. They have reduced their physical activities which has increased many health conditions. The mental health of these students has also been affected greatly. People in their teens and early twenties have shown a sign of stress along with other mental health disorders. This can also lead to serious health conditions such as schizophrenia, insomnia, or anorexia. Students with stress of online school due to screen time can be reduced by peer counseling, recreational activities, or physical workouts. Schools should also take measures in helping the students relieve stress by reducing school hours or having more clubs/activities during the week.

Acknowledgements

I am grateful to my parents for their endless support during this work. I would also like to thank Mr. Chandan Singh for his guidance. Additionally, I would like to extend my thanks to IJHSR team. I thank the respondents of the survey. Without their help, this work would not have been completed.

References

1. Acter T; Uddin N; Das J. Evolution of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) as Coronavirus Disease 2019 (COVID-19) Pandemic: A global Health Emergency. Sci Total Environ Journal **2020**, 730, 138996.
2. Hossain MM; Tasnim S; Sultana. A Epidemiology of Mental Health Problems in COVID-19: A Review F1000Res. **2020**, 9, 636.
3. Altmann DM; Douek DC; Boyton RJ. What Policy Makers Need to Know about COVID-19 Protective Immunity. Lancet. **2020**, 395(10236), 1527–1529.
4. Kuguyo O; Kengne AP; Dandara C. Singapore COVID-19 Pandemic Response as a Successful Model Framework for Low-Resource Health Care Settings in Africa? OMICS. **2020**, 24(8), 470–478.
5. McKee M: Learning from success. How has Hungary Responded to the COVID Pandemic? GeroScience. **2020**, 42(5), 1217–1219.
6. Esposito S; Principi N; Leung CC. Universal Use of Face Masks for Success Against COVID-19: Evidence and Implications for Prevention Policies. European Respiratory Journal. **2020**, 55(6), 2001260.
7. Robbins T; Hudson S; Ray P. COVID-19: A New Digital Dawn?. Digit Health. **2020**, 6, 2055207620920083.
8. Ting DSW; Carin L; Dzau V. Digital Technology and COVID-19. Nat Med. **2020**, 26(4), 459–461.

DOI: 10.36838/v4i4.14

9. Király O; Potenza MN; Stein DJ. Preventing Problematic Internet Use During the COVID-19 Pandemic: Consensus Guidance. Comprehensive Psychiatry. **2020**, 100, 152180.
10. Wang X; Li Y, Fan H. The associations Between Screen Time-Based Sedentary Behavior and Depression: A Systematic Review and Meta-Analysis. BMC Public Health. **2019**, 19(1), 1524. .
11. Janssen X; Martin A; Hughes AR. Associations of Screen Time, Sedentary Time and Physical Activity with Sleep in Under 5s: A Systematic Review and Meta-Analysis. Sleep Med Revieew. **2020**, 49, 101226.
12. Kuss DJ; Griths MD. Internet and Gaming Addiction: A System atic Literature Review of Neuro Imaging Studies. Brain Science. **2012**, 2(3), 347–74.
13. Lanca C; Saw SM. The Association Between Digital Screen Time and Myopia: A Systematic Review. Ophthalmic Physiology Optics. **2020**, 40(2), 216–229
14. Acharya JP; Acharya I; Waghrey D. A Study on Some of the Com mon Health Effects of Cell-Phones amongst College Students. Jounal of Community Medical Health Education. **2013**, 3, 214.
15. Maras D; Flament MF; Murray M. Screen Time is Associated with Depression and Anxiety in Canadian Youth. Prev Med. **2015**, 73,133-138.
16. Boers E; Afzali MH; Newton N; Conrod P. Association of Screen Time and Depression in Adolescence. JAMA Pediatr. **2019**, 173(9), 853-859.
17. Cellini N; Canale N; Mioni G; Costa S. Changes in Sleep Pat tern, Sense of Time and Digital Media use During COVID-19 Lockdown in Italy. Journal of Sleep Research. **2020**, 29, e13074.
18. Melissa Han. Emotion and Consumption Profiles in a COVID-19 Environment. International Journal of High School Research. **2021**, 3(1), 30-37.

Author

Laisha Gulia is a high school student at Good Shepherd International School in Ooty, India. She enjoys studying psychology and economics. She plans to do undergraduate work in behavioral economics. She enjoys music and playing the guitar.

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

The Benefits of Using Indirect Liquefaction Including Fischer-Tropsch Process for the Automotive Industry

Özge Dinç, Güney Baver Gürbüz
Üsküdar American Academy, Selamiali, Vakıf St. No:1, İstanbul, 34664, Turkey; ozgedinc2004@gmail.com

ABSTRACT: This paper investigates the steps of indirect liquefaction, specifically on the Fischer-Tropsch process. In particular, the essence of Fischer-Tropsch is elaborated, including its history and prerequisites. In the Fischer-Tropsch process, the catalytic conversion of synthetic gas produced from carbon feedstocks to synthetic fuels, works in specific conditions of temperature and pressure. This paper investigates the benefits of the diesel fuels derived from Fischer-Tropsch processes to the automotive industry, including the relatively low cost, the low greenhouse gas emission rates, and the facilitated process of gas conversion to synthetic fuel. In this paper, current and past investigations and experiments have been examined. As a result, several advantages of synthetic fuel over conventional crude oil were determined.

KEYWORDS: Energy: Chemical; Alternative Fuels; Automotive Industry; Indirect Liquefaction; Fischer-Tropsch; Synthetic Fuel.

Introduction

Nowadays, with the globally increasing energy market and the need for energy preservation, many kinds of energy and fuel are being sought. Especially for the industrial areas in a search for an extensive amount of fuel, the application of the Fischer Tropsch (FT) process can be a leading energy source. The FT process, including gas-to-liquid conversion, has the potential to produce high-value automotive fuels and petrochemicals from fossil and renewable sources. However, FT fuels currently are not a significant fraction of the global diesel fuel market due to the large capital investment required for production.

This research paper will investigate the benefits of using synthetic fuels produced via Fischer-Tropsch synthesis rather than conventional crude oil in the transportation industry and elaborate on the specific reasons why the merits of the mechanism outweigh the drawbacks. The paper starts with an introduction to indirect liquefaction and goes into the background, chemistry, and application mechanisms of several steps of conversion: feed-to-syngas conversion, preparation of syngas, facilities for the FT process, the Fischer-Tropsch process itself, and syncrude-to-products conversion. While alluding to the use of Fischer-Tropsch mechanisms in the automotive industry throughout the paper, there is a separate section merely mentioning the uses and benefits of FT-derived synthetic fuels in automobiles.

Discussion

Indirect Liquefaction:

Indirect liquefaction is the conversion of a feed, in this case, carbon-based energy sources, into a useful chemical or liquid fuel through intermediate steps where the feed is first converted into synthetic gas (syngas), then into synthetic fuel (syncrude), and finally into useful products. The liquefaction technology is generally defined as X-to-liquid (XTL) and more commonly classified as coal-to-liquid (CTL), gas-to-liquid (GTL), or biomass-to-liquid (BTL) conversion depending on the primary source of the syngas.

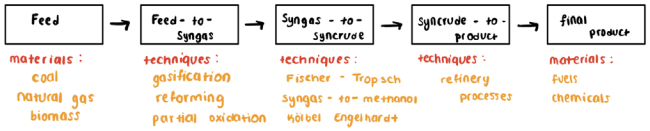


Figure 1: Feed-to-liquids (XTL) conversion and its overall indirect liquefaction process.

Feed to Syngas Conversion:

All the raw feed materials are already costly in terms of logistic commerce, and their requirement for some form of feed pretreatment before they are completely suitable for the conversion process also adds up to the cost. Because the carbon content of biomass is not concentrated at a single point of origin, it has a low energy density. As a result, the feed logistics involved in collecting and transporting biomass from its source to the indirect liquefaction facility add significantly to the cost and complexity of the process. Although it is already available from a pipeline supply, the widely used natural gas is hard to transport in gaseous form, and hence must be arranged into a condensed form. However, the ease provided by the raw feed materials for the next steps of the liquefaction makes it worth spending a large sum of money.[1]

The high cost of the feed pretreatment makes it critical for facilities to select the syngas production technology carefully and accordingly. It should be compatible with the feed, and it should ideally be selected to meet the syngas component requirements of the syngas-to-syncrude conversion technology.[1] There are three catalyst-involving ways in which the feed materials can be converted into syngas, a mixture of carbon monoxide (CO) and hydrogen (H_2): steam reforming, adiabatic oxidative reforming, or gasification.

DOI: 10.36838/v4i4.15

Method 1: Steam Reforming:

Steam reforming is usually used when the feed is natural gas.[1] A steam reformer is a reactor consisting of a fired heater. The heater will be filled with waste fuel coming from the combustible materials found in the waste coming from the households or industry. Then, the fuel will be burnt to supply the heat needed for reforming. A light hydrocarbon feedstock, primarily natural gas, is reacted with steam at extreme temperatures and pressures in nickel-based catalyst-filled tubes to create synthesis gas in the steam reforming process. Hydrogen and carbon monoxide make up the majority of this gas, but other gases including carbon dioxide and nitrogen, as well as water vapor, are also present. The produced syngas will have a high H_2:CO ratio since natural gas mostly consists of hydrocarbons and steam (H_2O).

Method 2: Adiabatic Oxidative Reforming:

Adiabatic oxidative reforming is mostly used when the feed is methane-rich and oxidant-consisting: biomass material or, in some cases, natural gas.[1] The heat needed for reforming is directly supplied by the combustion of part of the feed. The carbon monoxide produced after combustion then combines with water to generate carbon dioxide and additional hydrogen in a water-gas shift reaction. The process of partial oxidation is exothermic, meaning it generates heat. In most cases, the process is much faster than steam reforming and requires a smaller reactor vessel. Although this allows for a more compact design than a steam reformer, it has the disadvantage of requiring an associated air separation unit due to the combustion of material consisting of oxygen. The produced syngas will have a moderate H_2:CO ratio.

Method 3: Gasification:

Solid feed materials such as coal have to be gasified to produce proper syngas. Gasification is the conversion of organic or fossil-based carbonaceous resources with a controlled amount of oxygen into syngas and CO_2. It can take place at different temperatures.

However, the high-temperature conditions are the most advantageous ones, as they not only produce the required syngas but also clean them from pyrolysis products, which are organic materials heated in the absence of oxygen.[1] This helps the facility by simplifying the downstream gas cleanup and refining the process in further steps.

Syngas Cleaning, Cooling, and Conditioning:

The produced syngas via steam reforming, adiabatic oxidative reforming, or gasification includes compounds that are referred to as FT poisons. Poisons are formed from the presence of heteroatoms, any atomic element other than C and H, and are most commonly converted into gaseous compounds like H_2S, COS, and NH_3. These are known to disrupt the proceeding of the FT process by disabling the related catalysts, mainly Fe and Co. Hence, syngas cleaning, cooling, and conditioning through industrial applications are required to remove the poisons and ensure the continuity of the process.[1]

Although many nitrogen-containing compounds or bromides are also considered poisonous for FT, the universal poison for FT catalysts is sulfur.[1] Both iron (Fe) and cobalt (Co) catalysts are deactivated in the presence of sulfur; however, Co is more sensitive to it. This is one reason why the facilities would prefer cobalt catalysts for FT processes with syngas derived from natural gas, which turns out to be relatively lower in sulfur content with a higher H_2:CO ratio.[1]

Through syngas cooling, nitrogen compounds and pyrolysis products can be removed from the syngas. Nitrogen compounds are mostly soluble in an aqueous medium, so it is almost like using water-washing to remove NH_3. However, the pyrolysis products need some more developed industrial mechanisms to be separated from the syngas.[1]

Syngas conditioning is the last step before the FT process. The H_2:CO ratio is adjusted to meet the specific requirements of the chosen FT technology mainly by water gas shift conversion :

$$(n)\ CO + (n)\ H_2O \rightarrow (n)\ CO_2 + (n)\ H_2$$

The syngas can be turned into syncrude once the feed-to-syngas conversion is complete, and the syngas has been cleaned and conditioned using the appropriate technological combinations that are available.[1]

Fischer-Tropsch Requirements:

The process of the conversion of syngas into syncrude can take place in a myriad of ways, one being the Kölber Engelhardt process, but in this case, the main focus will mainly be on the Fischer-Tropsch process.

In the Fischer-Tropsch process, CO and H_2- syngas- goes under a catalytic chemical reaction and are converted into primary hydrocarbons- syncrude- of various weights according to the following chemical equation:

$$(2n + 1)\ H_2 + (n)\ CO \rightarrow C_nH_{(2n+2)} + (n)\ H_2O$$

The FT procedure conditions should be chosen to maximize the formation of higher molecular weight hydrocarbon liquid fuels, possessing higher value due to the high energy that will be produced when broken down.[2] Hence, the type of catalyst, the type of reactor, temperature, and pressure conditions should be inclusively considered. The facility should go with the combination that will provide the highest desired product yield. One can select the reactor, temperature, catalyst, and pressure conditions to produce a specific product.

Catalysts:

In the 1920s, Franz Fischer and Hans Tropsch discovered that they could produce paraffin hydrocarbons by passing the syngas over catalysts in optimum temperature and pressure conditions.[3] They worked with many different metal catalysts, but it was found that the most beneficial options were Fe and Co in the end.

Fischer and Tropsch considered the conditions of the FT process and the possible outcomes, and they concluded that the ideal catalyst should both adsorb CO, preferably in a dissociative way, and H_2. Also, the metal oxide of the catalyst, co-produced as a result of the dissociative adsorption of CO or reaction of the metal with the co-produced water, should easily be removed.

Firstly, they worked with early transition metals. These elements successfully dissociated the CO molecules but did not adsorb H_2. Furthermore, their oxides were not the kinds that could be reduced under usual FT conditions. The final result was that these molecules were not the optimum FT catalysts.

Later on, they worked with late transition metals and Group 12 elements such as Zn, Cd, and Hg.[4] These elements showed either no CO adsorption or adsorbed it in a non-dissociative way. Although they adsorbed H_2 efficiently and produced metal oxides that were easily reduced, they were also not the ideal FT catalysts.

They then worked on the elements Ni, Fe, Co, and Ru. Nickel was not a common option, since it promoted the formation of an undesirable byproduct: methane. Ruthenium was not available either because it was too rare and expensive to be used on a huge scale. In the end, Fischer and Tropsch were left with Fe and Co. Since both elements have different properties, they promoted different mechanisms for FT.

Iron Catalysts:

Iron catalysts are relatively low-cost and have a higher water-gas shift activity. Therefore, they are more suitable for lower H_2:CO ratio syngas. So, they are widely preferred for processing syngas derived from coal gasification or biomass reforming.[2] One advantage of iron catalysts is that they can be operated both in high temperatures between 300 to 350 °C and in low temperatures between 220 to 270 °C.

When iron is utilized at high temperatures, the primary products are composed of gasoline, mainly 2,2,4-trimethylpentane (iso-octane) with a chemical formula of C_8H_{18}. Furthermore, the high-temperature conditions will trigger the formation of secondary and undesired products, such as ketones, that will need further removal operations.[2] However, at low temperatures, diesel production is more dominant with higher carbon number products and nearly no byproducts. Due to this fact, if the automotive industry were to able, they would choose and use low temperatures with an iron catalyst. This will cost less both for the catalyst & operation and will be easier to obtain. Still, iron is not the primary preference of the transportation sector.

Iron also has an advantage over cobalt in terms of being open for promoters. Alkali promoters, such as K_2O, can be reacted with iron before the process to further enhance the efficiency and rate of CO dissociation. Promoters operate by interacting with active components of catalysts to enhance their catalytic activity.

Cobalt Catalysts:

Cobalt catalysts are much more expensive than iron catalysts and they only can operate in low-temperature regimes. On the other hand, being active in lower synthesis pressures and temperatures significantly decreases the operating cost. Furthermore, they have a longer lifetime when compared to iron when the catalyst surface area is maximized. Maximizing efficiency can be done by dispersing cobalt onto aluminium oxide (Al_2O_3), silicon dioxide (SiO_2), or titanium dioxide (TiO_2).[5] The catalyst will need replacement less frequently. These advantages offset the high catalyst cost. Thus, cobalt becomes a much more useful alternative for FT catalysis.

Cobalt is not as active as iron when it comes to catalyzing the water-gas shift reaction. It is mostly preferred when the feed material is a natural gas (GTL Conversion). The automotive industry's primary feed material is mainly natural gas due to the produced cetanes and non-produced sulfur content. Hence, the industry usually makes use of cobalt catalysts rather than iron ones.

Similar to the iron catalysts, the main product after cobalt-based low-temperature FT is diesel. This is another advantage of cobalt for the automotive industry because their desired product is diesel molecules, and the operation cost for such a product will be lower when compared to higher temperature conditions.

Table 1: Syncrude compositions that are typical of iron-based high-temperature FT (Fe-HTFT), iron-based low-temperature FT (Fe-LTFT), and cobalt-based low-temperature FT (Co-LTFT) syntheses.[1]

Product Group	Carbon Range	Compound Class	Syncrude Composition (mass-%)		
			Fe-HTFT	Fe-LTFT	Co-LTFT
Tail Gas	$C_1 - C_2$	Alkane Alkene	17.2 5.6	5.3 1.0	6.6 0.1
LPG	$C_3 - C_4$	Alkane Alkene	3.0 21.2	1.8 6.0	1.8 3.4
Naphtha	$C_5 - C_{10}$	Alkane Alkene Aromatic Oxygenate	4.3 25.8 1.7 1.6	3.3 7.7 0 1.3	12.0 7.8 0 0.2
Distillate	$C_{11} - C_{22}$	Alkane Alkene Aromatic Oxygenate	0.9 4.8 0.8 0.5	13.5 5.7 0 0.3	20.8 1.1 0 0
Residue / wax	C_{22+}	Alkane Alkene Aromatic Oxygenate	0.4 1.6 0.7 0.2	49.2 0.7 0 0	44.6 0 0 0
Aqueous Product	$C_1 - C_5$	Alcohol Carbonyl Carboxylic Acid	4.5 3.9 1.3	3.9 0 0.3	1.4 0 0.2

Temperature:

The temperature conditions of FT vary accordingly with the relative usage of catalysts. This temperature moderation can be a significant factor in the composition of the syncrude. The temperatures between 180 °C - 300 °C are set as ideal for the process to take place.

When gasoline is the desired product, facilities mostly prefer high-temperature conditions because gasoline is a low-carbonated and highly hydrogenated liquid, in line with the typical outcomes of high-temperature catalysis. High-temperature FT also yields low molecular weight alkenes, which are classified as undesired byproducts in most cases for the further processing of the fuels formed through FT.[2] High-temperature conditions will cost higher for the facility since it is hard to initiate and maintain such temperatures within a mechanism.

As it is desired in the automobile industry case, low-temperature FT processes yield high molecular weight diesel molecules. They also do yield high molecular weight linear waxes. Low temperatures are easier to maintain because most industrial facilities have already developed systems suitable for such conditions.

The regulation of the temperature conditions should be made in accordance with the desired products primarily. The facilities can then consider the operation costs in specific conditions as a secondary factor. This is crucial in this matter.

DOI: 10.36838/v4i4.15

Table 2: Comparison between the product compositions obtained from FT synthesis.[1]

Compound Class	HTFT	LTFT
Alkanes (Paraffins)	> 10 %	major Product
Cyclo-alkanes (napthenes)	< 1 %	< 1 %
Alkenes (olefins)	major product	> 10 %
Aromatics	5-10 %	< 1 %
Oxygenates	5-15 %	5-15 %
Sulfur Compounds	0 %	0 %
Nitrogen Compounds	0 %	0 %
Organometallics	Carboxylates	Carboxylates
Water	major by-product	major by-product

Pressure:

The ideal pressure for the process is determined to be between 20-50 bars. High pressures support the production of waxes, while moderate pressures are suitable for diesel production and more favorable for the automotive industry.

Reactors:

The FT process is a reaction with a heat outcome. This exothermic reaction needs heat removal for the continuity of the process in the design of a commercial reactor. In general, there are three different types of reactors: fixed bed, fluidized bed, and slurry bed.

Fixed bed reactors are used to produce high carbon number liquid hydrocarbons and waxes. This could be favorable in normal conditions for the automotive industry. However, the automotive industry doesn't only consider the products, but also the process. Hence, slurry bed reactors that offer better temperature control and higher conversion efficiency than fixed bed reactors are preferred in automotive industrial processes. On the other hand, fluidized bed reactors are mostly used to produce low molecular weight hydrocarbons like gasoline.

Fischer-Tropsch Process:

First carried out by German chemists Franz Fischer and Hans Tropsch in the 1920s, the Fischer-Tropsch (FT) process is a catalytic chemical reaction that allows carbon-based energy sources to be converted into products such as fuels or chemicals. Back then, Fischer, the director of the Kaiser-Wilhelm Institute for Coal Research in Germany, wanted to perform research on coal. It was one of the most used fuels back in the 1920s, and therefore, every country needed to know about the processing of coal for the further development of industrial factories and the weapons industry. However, Fischer soon realized that Germany did not have enough oil resources to combine with coal and make an efficient system of power. Hence, he shifted his goal towards converting coal to petroleum, now known as the FT process. Fischer worked with Hans Tropsch, another chemist working in the Kaiser-Wilhelm Institute, on reducing CO by using excess H_2. They spent nearly a decade working on the process, but their efforts didn't pay off until they replaced the chemical catalysts with cobalt and iron derivatives. By 1944, nine commercial-sized Fischer-Tropsch stations were under construction with a production capacity of 4.1 million barrels of hydrocarbon fuels annually. These stations provided nearly 10% of German fuels during World War II, hence causing controversy between countries, as only Germany knew about the usage of the system. Nonetheless, in the past, it was an undeniable fact that it was a revolutionary finding in industrial fuel production.

Today, as soon as all the prior conditions of the required FT technology are set, the process can take place through a variety of steps with intermediate products.

The FT reaction is initiated by the adsorption of the reagents on the catalyst surface. The catalysts work on the syngas to convert its components CO and H_2 into long-chain paraffin. The C-O and H-H bonds are first cleaved by the catalysts.

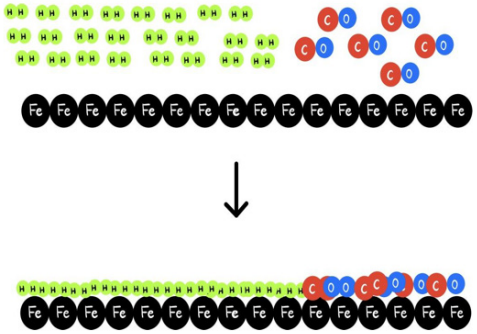


Figure 2: The catalysis of CO and H_2 over Fe catalyst.

Then, the separated hydrogen atoms contribute to the formation of hydrocarbons and water by forming new covalent bonds with separated carbons and oxygen, respectively.

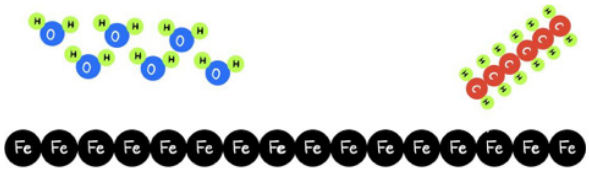


Figure 3: The formation of hydrocarbons and water as a result of syngas catalysis.

These reactions occur in the reactor section of the FT mechanism. Furthermore, during this chain of reaction, heat is released. The outcoming heat is removed by a cooling tube mechanism inside the reactor by generating high-pressure steam.

The formed H_2O and hydrocarbons are then transferred into a separator, where the mixture is condensed, and water is removed. This both aids in the further cleaning of the main product and the definite formation of liquid hydrocarbons. In the separator, there could be formation of light gases as byproducts. These will be sent to the initial mechanism of feed transformation for further processing in possible future operations.

Moving through with the FT process, the condensed hydrocarbons are sent to a hydrocracker. In the hydrocracker, the very long-chain waxy hydrocarbons are heated and reacted with hydrogen to form high-quality fuels. This step is the most essential step in the overall process for the automobile industry. Facilities can only reach their desired product, diesel, only by this last reforming process on the long-chain paraffins.

Synthetic Fuel to Final Product:

The products of the Fischer-Tropsch reaction are not directly used in industrial processes, as the byproducts formed as a result of a catalyst reaction will reduce the efficiency of the synthetic fuel. Synthetic fuel can be transformed into final and useful products only if it goes under the upgrading and refining processes. In refinement processes, the aim is to remove impurities or unwanted elements from the synthetic fuel. The experts describe this step as an inseparable part of the Fischer Tropsch conversion process.

Upgrading is the primary step of further enhancement of the synthetic fuel where some other intermediate products are formed. The produced chemicals are easier to process through an industrial mechanism. Once the syncrude is upgraded, it can either be partially refined or categorized as stand-alone.

When partial refining takes place, some of the syncrude is refined to the final products, whereas some are destined for blending with coal liquid for further use as a final product.[1] Nevertheless, in stand-alone refining, all the syncrude is directly converted into final products. This requires some special and complex mechanisms. Hence, many Fischer-Tropsch entrepreneurs have built stand-alone refineries within the facility.

Automobile Industry:

Automotives are one of the most energy-using machines that are being manufactured, even one single automobile wastes tons of liters of fuel. Today's 97% of energy demands of are met by petroleum-based fuels.[7] While this extensive use of fossil fuels creates concerns for the environment, the high costs of petroleum fuels are also to the detriment of the industry. Hence, in the case of diesel fuel for automotive engines, many alternative fuel sources are being discussed as plausibly viable through the petrochemical industry. Fischer-Tropsch-based diesel produced from the syngas is one commercially viable alternative for this energy demand. In other words, the hydrocarbons produced by the FT process can be refined and used in place of more conventional liquid fuels derived from crude oil.[6] Since these hydrocarbons can be used in automotive engines directly and are almost identical to fuels refined from crude oil, they are set apart from most currently available fuels, such as ethanol, that have to be mixed with gas, and require special engines, or create challenges for low-temperature operations.[8]

As mentioned above, the desired fuel for automobiles is most commonly diesel, which is the main product of the indirect liquefaction of natural gas. The synthetically produced diesel fuel after gas-to-liquid conversion will have a high cetane number and no sulfur content. This reduces the rate of residuals, such as minerals in the process, facilitating the derivation of diesel directly from the FT mechanism.[5] In addition to the facilitated processing, natural gas is an available source all over the world and has a low cost compared to high petroleum prices. Many automobile facilities would prefer this less complex and less costly way of obtaining the main product over the more difficult and more costly mechanisms of crude oil processing.

Furthermore, the main point of using syncrude instead of petroleum is to substitute fossil fuels with renewable feedstocks and make use of them through an environmentally friendly mechanism. The global climate concerns will be eradicated over time by using FT fuels to help with the flaring of natural gas and emission of a tremendous amount of CO_2, each time fossil fuels are used directly. Christodoulos Floudas, a professor of chemical and biological engineering at Princeton University, claims that even if a country immediately converted to zero-emitting electric or fuel cell vehicles, millions of internal combustion vehicles would still be used.[12] According to him and his research data, switching to synthetic fuels, nevertheless, could help the countries reduce CO_2 emissions at a high rate. The heavy metal and sulfur contaminants of petroleum fuels can be captured in synthetic fuels before they are shipped out. In regards to these environmental benefits, the American Institute of Chemical Engineers (AIChE) calls for greater integration of energy sources and urges policymakers of every country to consider chemical conversion processes as a potential method to produce cleaner and cheaper fuels.[8]

Expectedly, as the Fischer-Tropsch facilities have been refined to increase efficiency over years, their cost of the building has increased significantly.[8] For example, it is estimated that roughly 45 billion dollars will be required for the entire system that is expected to be composed of around 100 facilities all around the United States.[8] This might seem costly, but Richard Baliban, a chemical and biological engineering graduate from Princeton in 2012, claimed that as long as crude oil is between $60 and $100 per barrel and it continues increasing at the current rate, synthetic fuels are very competitive and can be profitable over time.[8]

Conclusion

Low greenhouse gas emissions, a streamlined procedure, and reduced cost are just a few of the benefits of synthetic fuel over conventional crude oil that make it a favorable alternative for consideration as a fuel source for the automotive industry. Synthetic fuels would enable carbon reduction with the present fleet of cars on the road. Millions of internal combustion vehicles would still be on the road even if the country switched to zero-emission electric or fuel cell vehicles right away. By converting to synthetic fuels, the governments would be able to

DOI: 10.36838/v4i4.15

minimize emissions, even if they could not be totally eradicated. Synthetic fuels are cleaner in many aspects than petroleum fuels. Petroleum fuel impurities such as heavy metals and sulfur can be collected in synthetic plants before the fuel is transported out. Unlike many biofuels, synthetic fuels can be utilized in gasoline and diesel engines without modification and hence are more practical for use in industrial applications. In terms of the investment in the Fischer-Tropsch process, the cost of creating the facilities would be the largest contribution to the price of synthetic fuel, followed by the purchase of biomass and finally power. The cost could be much lower if plants didn't use biomass and instead ran on coal and natural gas, but most of the environmental benefits then would be lost.

Currently, there are a few facilities in Sasolburg and Malaysia that are operating and a few under construction in Nigeria, Qatar, China, and the United States. Applications to daily life have already been experienced by Syntroleum and Audi. Syntroleum, a publicly traded American corporation, has produced over 400,000 gallons of diesel and jet fuel at its demonstration plant outside Tulsa, Oklahoma, utilizing the Fischer–Tropsch process using natural gas and coal. Syntroleum is developing coal-to-liquid and gas-to-liquid plants in the United States, China, and Germany, as well as gas-to-liquid plants around the world, to market its licensed Fischer–Tropsch technology. The ultra-clean, low sulfur fuel, which uses natural gas as a feedstock, has been extensively studied by the US Department of Energy (DOE) and the US Department of Transportation (DOT). Syntroleum has recently been collaborating with the US Air Force on the development of a synthetic jet fuel blend that will assist the Air Force to lessen its reliance on imported petroleum. A B-52 took off for the first time from Edwards Air Force Base, California, on December 15, 2006, propelled only by a 50–50 blend of JP-8 and Syntroleum's FT fuel. The flight test, which lasted seven hours, was deemed a success. In 2007, the test program came to an end. This initiative is part of the Defense Department's Assured Fuel Initiative, which aims to establish secure domestic energy sources for the military. The Pentagon wants to cut down on its reliance on foreign crude oil and get around half of its aircraft fuel from alternate sources by the near future.[9] In partnership with Sunfire, Audi produces E-diesel on a small scale with two steps, the second one being FT.

All these applications and plans require great optimization and efficiency. To that end, the increasing number of enhanced Fischer-Tropsch facilities will utilize synthetic fuels more than ever for the automotive industry while promising great potential in other areas as well. The works illustrated herein help provide a better understanding of the indirect liquefaction process, including Fischer-Tropsch, and highlight its benefits primarily for the automotive industry and our environment.

References

1. Klerk, Arno de. Fischer-Tropsch Refining. Weinheim, Germany: Wiley-VCH, 2011. Digital file.
2. "10.2 Fischer Tropsch Synthesis." National Energy Technology Laboratory. Accessed July 4, 2021.https://www.netl.doe.gov/research/coal/energy-systems/gasification/gasifipedia/ftsynthesis.
3. Milmo, Sean. "Go faster FT catalysts." Where Science Meets Business. Last modified 2011. https://www.soci.org/chemistry-and-industry/cni-data/2011/9/go-faster-ft-catalysts.
4. Aresta, Michele, Antonella Colucci, Angela Dibenedetto, Franck Dumeignil, Gennaro Agrimi, Efthymia Alexopoulou, Elisavet D. Bartzoka, Antonio Buonerba, Fabrizio Cavani, and Myrsini Christou. Biorefineries - an Introduction. Berlin, Germany: De Gruyter., 2015. Digital file.
5. "Fischer-Tropsch."Science Direct. Last modified 2015. https://www.sciencedirect.com/topics/engineering/fischer-tropsch.
6. "Fischer-Tropsch Process." Biofuels Academy. http://biofuelsacademy.org/index.html%3Fp=388.html.
7. Schaub, G., M. Rohde, and A. Mena Subiranas. "Fischer-Tropsch Synthesis - Development and Perspectives." DGMK/SCI-Conference. Last modified October 4, 2006. http://www.osti.gov/etdeweb/servlets/purl/20840756.
8. "8.5 Fischer-Tropsch Process to Generate Liquid Fuels." Penn State College of Earth and Mineral Sciences. https://www.e-education.psu.edu/egee439/node/679.
9. Zamorano, Marti (2006-12-22). "B-52 synthetic fuel testing: Center commander pilots first Air Force B-52 flight using solely synthetic fuel blend in all eight engines". Aerotech News and Review.
10. Dodaro, John. "Fischer-Tropsch Process." Stanford University PH240. Last modified December 11, 2015. http://large.stanford.edu/courses/2015/ph240/dodaro1/#:~:text=The%20Fischer%2DTropsch%20 process%20is,means%20of%20a%20metal%20catalyst.
11. Zhao, Jianshe, Ying Zheng, Tom Whidden, Haiping Zhnag, and Zhen Dong. "Highly Selective Fischer-Tropsch Synthesis for C10-C20 Diesel Fuel under Low Pressure." Research Gate Last modified February 2017. https://www.researchgate.net/publication/313729310_Highly_selective_Fischer-Tropsch_synthesis_for_C_10_-C_20_diesel_fuel_under_low_pressure
12. Sullivan, John. "Synthetic fuels could eliminate entire U.S. need for crude oil, create 'new economy." Princeton University. Last modified November 27, 2012. https://www.princeton.edu/news/2012/11/27/synthetic-fuels-could-eliminate-entire-us-need-crud e-oil-create-new-economy.
13. Bowen, Brian H., and Marty W. Irwin. "Coal Gasification & Fischer-Tropsch." Indiana Center for Coal Technology Research. Last modified July 2006. http://www.mariloureiro.net/ciencia/Gasificacao/Coal-To-Liquids.pdf

Author

Güney Gürbüz from Turkey was born in 2004. After starting high school at Uskudar American Academy, his interest in sciences and engineering has increased. In 2020, he created his own website with his friend, in which he publishes articles concerning the integration of sciences in modern engineering. He also established an academic research conference for high school students.

Özge Dinç was born in 2004. Her career plans include chemical sciences. At Uskudar American Academy, by taking high level science classes, her interest has increased, and she has been working on several projects like competitions to make it possible for every high school student to experience scientific research.

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Contouring a User Centered Chatbot for Diabetes Mellitus

Harnishya Palanichamy
Hebron School, Lushington Campus, Ootacamund, Nilgris District, Tamil Nadu, 643001, India; harnishya.bts@yahoo.com

ABSTRACT: Diabetes Mellitus (DM) is a chronic disease. Its management requires continuous adherence to medical care, self-management, and monitoring, to mitigate the risks. During COVID, the difficulties faced in accessing health facilities, created a sense of disconnect between doctors and patients. A Chatbot for Diabetes Mellitus, DBOT, was designed using a set of questions grouped under components- Diabetes History, Medical History, Family History, Medical History (for women only), Eating/Exercising Habits, Tracking Health Status (Lab Test Results), Prescription Status, Knowledge about reducing Risks, Frequently Asked Questions (FAQ), and DBOT Usage. To design the DBOT, JAVA Coding, xml designing language and an Android Studio tool are used. The flow charts and coding sheets for the components were devised. The DBOT was shared with the diabetic patients for their utilization and through a questionnaire survey and the feedback from the respondents about user experience, were collected. The questionnaire data were analyzed using statistical techniques like Two Sample 't" Test, Chi square, One-Way ANOVA, Correlation, and Factor Analysis to determine comparisons, relations, and associations between the components of DBOT. The future steps include incorporating the proposal, recommendations and suggestions of the users and restructure the DBOT with enhanced effectiveness.

KEYWORDS: Systems Software; Mobile Apps; Diabetes Mellitus; DBOT; Questionnaire; Statistical techniques.

Introduction

Health care is the prevention, treatment, complications, management of illness and the preservation of mental and physical well-being through the services offered by the medical, nursing, and allied health professions.[1] The objective of a health care system is to strengthen the health of the people in a society in an effective method. A health care provider, principally a physician, always possesses the potential to elucidate, expound, and supervise the health condition of the patients. A robust interaction and exchange of information between the doctor and patients is a major requisite that develops trust between participants and in succession, reduces the medical risks.

Diabetes Mellitus (DM) is a chronic disease associated with greater rates of cardiovascular problems, kidney disease, vision problems, and non-traumatic amputations.[2] DM cannot be cured, but it can be detected, managed,[3] and its prevention through continuous lifestyle monitoring can delay the further development of the disease.[4] DM management is long, costly, and requires continuous adherence to medical care and in addition, it requires ongoing self-management and monitoring to mitigate the potential risks.[5] The regular and daily decisions made by patients with diabetes (like, eating healthy foods, tracking physical activity, administering insulin and other medications, monitoring blood glucose, undergoing foot and eye care, participating in laboratory studies, making regular clinic visits, maintaining health education) are very important for DM management.[6] Poor adherence to these activities can lead to significant mortality and morbidity, as well as poor quality of life.[7,8]

In general, people with diabetes are more likely to have severe symptoms and complications when infected with any virus.[9] During COVID, the difficulties faced in accessing health facilities, created a sense of disconnect between doctors and the patients.[10] In particular, for diabetes patients, as there is a possibility of disruption to the routine lifestyle management measures like physical activity and psychological stress management due to lockdown, people with diabetes must explore ways to manage their diet, exercise and stress in consultation with the treating physician to prevent worsening of symptoms.[11] Older adults with diabetes, during COVID, are at highest risk of adverse outcomes and mortality caused by the virus and this has greatly affected by their inability to access and receive health care, obtain diabetes medications and supplies, and maintain a healthy lifestyle and social connections.[12]

As a solution, information and communications technologies (ICT) can assist both patients and physicians to improve bonding by introducing them to electronic health (e-Health).[13] According to the World Health Organization (WHO), mobile health (MH) is a component of e-Health and by using mobile phones and smart devices, MH provides promising opportunities to improve diabetes prevention, detection, and self-management with continuous measurements of a patient's bio-signs.[3] In other words, MH supports the transition from clinic-centric to patient-centric healthcare where each agent-hospital, patient, physician, and service are seamlessly connected to each other.[14] The need for a reliable and accurate diagnosis awakens the rise of a new generation of healthcare technology, the Medical Chatbot.[15]

Medical chatbots are a technology that makes interaction between man and machine possible by using natural language processing with the support of dialog flow, because chatbots are reliable, compatible, and provide instant replies.[16] The rapid evolution of Medical Chatbots has opened a niche for Doctor-Patient communication, that minimizes the costs

DOI: 10.36838/v4i4.16

and time commitment on routine operations.[17] If a patient rushes in with an emergency, where every second is vital, the doctor can get the patient's information from previous records, related to other diseases, allergies, check-ups, etc., instantly using a medical chatbot.[18]

A medical chatbot paves the way for genetic diagnosis, clinical laboratory screening, and health communication[19] and the main idea of creating the chatbot is to replicate a human discussion.[20] As a tool with high utility among elderly and physically disabled people, a medical chatbot can help patients get solutions to all their health related issue at their fingertips.[21] Patients may also feel that the chatbots are safer interaction partners and hence patients disclose more medical information.[22] The medical chatbot supports sharing of real-time data between healthcare provider, physician, and patient and that provides end to-end comprehensive care and especially for the patients with diabetes, this real-time monitoring can avoid many adverse events.[23] And a chatbot will function as a virtual Diabetes physician to do a basic diagnosis on diabetic patients.[24] This healthcare chatbot system will help hospitals to provide healthcare support online 24 x 7, as it answers deep as well as general questions and serves many people at the same time with the same topic.[25]

My dad is a pioneer, qualified, experienced, and a certified Diabetologist, specialized in the research and treatment of all types of Diabetes, practicing in Tiruchirappalli City, Tamil Nadu state, India. He carries out thorough and complete clinical examinations. His treatment of diabetes includes the regular monitoring of blood pressure, blood glucose levels, insulin therapy, oral anti-diabetic medication, and the careful examination of heart, lungs, abdomen, the nervous system, and the feet. Through conversations with him, I garnered facts regarding the types, signs and symptoms, complications, diagnosis, treatment, and emergency management of diabetes. I was astonished to realize that all types of diabetes are controllable and must be managed for the rest of the person's life, either with insulin or oral anti-diabetic medicines. And those details and statistics have always fascinated me.

In a discussion with my dad, I became acquainted that the COVID pandemic forced the diabetes community to crucially understand the impact of the virus on diabetes patients, in recent days.

Among the diabetic patients, when the sugar level is well maintained and managed, the risk of acquiring illness due to COVID is minimal, even then, the mobility restrictions were an obstacle to access the doctor for check-ups during crisis times. I have noticed that my dad's patients residing in small townships, suburbs, and villages, located in different districts around my city, are unable to visit him to access their monthly or trimonthly follow-up, during the lockdown days.

Diabetic patients who are highly vulnerable/high-risk should be provided an opportunity to communicate with my dad and as a remedy, Whatsapp and Telegram groups were created for receiving the patient's medical and health queries. For some time, the group chats enabled the patients to seek consultation through posting queries. But eventually, maintaining different groups and sending messages resulted in time management and organization difficulties.

At this juncture, the situation inspired me to create a simple chatbot for my dad's patients. A chatbot should provide their users with a simple, valuable, reliable, and pleasing experience.[26] Similarly, I strongly felt that my chatbot should provide a valuable and a pleasing experience to the user and that subsequently will increase its dependability. Right at this point of time, I framed my research question as-

"How to design a user-friendly Medical Chatbot?"

I aimed to initiate a "Diabetic patients- Diabetologist" interlinkage for a finer comprehension of the basic details of the patients, with a simple chatbot. Though chatbots are available readymade, I wish to prepare myself for my dad's patient community. To start with, I referred articles about Diabetes to collect information about the Type1 and Type2 diabetes, classic symptoms, glucose level, blood sugar, Diabetic retinopathy, diabetic nephropathy, diabetic neuropathy, diabetic related foot problems, periodontitis, Diabetes Diagnostic criteria and management. Meanwhile I pursued the Coding skills and prepared a plan of action to design a chatbot, named DBOT, where D stands for Diabetes.

Methods

-A survey questionnaire, with 105 questions, was prepared and circulated to 100 respondents (diabetic patients), those included- Male (65 in number) and Female (35 in number), belonging to various age groups, living locality and educational qualifications. The questionnaire survey was done during September 2021. The questionnaire was sent to the respondents, who are the diabetic patients of my father. The respondents are chosen from the taluks of Tiruchirappalli district, namely- Lalgudi, Musiri, Manapparai, Manachanallur, Tiruchirappalli, Thottiyam, Thuraiyur, Tiruverumbur, and Srirangam.

-Statistical techniques like- Two Sample 't' Test, Chi square test, One Way ANOVA, Correlation and Factor Analysis are used to understand how easy the components are to use.

-Suggestions from the respondents, for contouring DBOT, as a User Friendly chatbot, were received and considered for upgrading the DBOT, in future.

Data Set:

To start with, I framed a set of questions related to Diabetes Mellitus and then grouped those questions under relevant components, namely- Diabetes History, Medical History, Family History, Medical History (for women only), Eating/ Exercising Habits, Tracking Health Status (Lab Test Results), Prescription Status, Knowledge about reducing the Risk, Frequently Asked Questions (FAQ), and DBOT Usage. To design DBOT, I used JAVA Coding language, xml designing language and Android Studio tool. The flow charts and Coding sheets for the components are devised.

After designing the DBOT, I determined to explore the effectiveness of the components included in the chatbot. For the same, I shared the chatbot with some of the active and diligent diabetic patients, to check its usage for a couple of weeks. I perceived that the maximum potential of chatbots will reach the patients, only when their challenges, needs, and expectations are met, because such an investigation determines the

variation in the users' individual perception and experience. To understand the user appropriation, satisfaction, perception, and preferences of DBOT, I preferred receiving the patient's feedback, which would help me to improve the effectiveness of my chatbot. The questionnaire feedback from the user will provide valuable guidance that might assist to mitigate the usability issues and strengthen the conversation effectiveness.[27] Similarly, I wish to scrutinize and examine the user experience, through a questions survey, to examine how convenient, comfortable, and feasible the DBOT is. Questionnaires are an important tool which provides patients the opportunity to voice their experience in a safe fashion and in turn, the health care providers gather information that cannot be picked up in a physical examination.[28] A multi-method approach with Pre- and Post-Test Questionnaire, user tests and short debriefing interviews were able to help me understand the user experience of any chatbot.[29] User feedback through the guided interviews and post-task surveys allow me to collect qualitative and quantitative data and help to discover the positive and negative aspects and develop personalized product experience for target user.[30]

Hence, I formulated a questionnaire related to the components and user experience of DBOT. The questionnaire had 105 questions and accepts answers in the form of the sentences and it has no restriction of the maximum limit of the words that can be entered. The questionnaire was disseminated to 100 diabetic patients. While choosing the respondents the significant factors considered are as follows.

1. Out of 100 respondents, 65 are Male and 35 are Female.
2. Age grouped under <15 (5 Male and 5 Female), 16-40 (20 Male and 10 Female), 41-60 (20 Male and 10 Female), and above 60 years (20 Male and 10 Female).
3. Locality- City, Town or Village.
4. Educational Qualification – Completed School or College Education or Illiterate.

The questionnaire comprises Yes/No questions and questions with 2,3, and 4 options as choices.

Data Processing:

The questionnaire was circulated to the diabetic patients, to record their feedback. Later the questionnaires were collected, and the data were tabled. The data were inspected using SPSS 16.0. The statistical techniques like Two Sample 't" Test, Chi square test, One Way ANOVA, Correlation and Factor Analysis are implemented, and the results are analyzed and interpreted. The results assisted in comparing the relationships between the components, in DBOT.

Results and Discussion

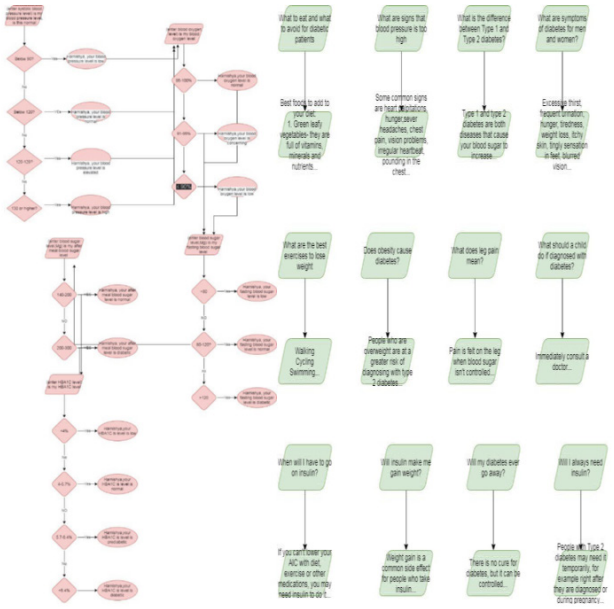
Designing DBOT :

The first step to designing my app was researching and collecting data from my dad, from answers to survey questions. The next step was creating a plan, which I did as a flow chart. The flowchart included all the data details the user could choose from, the answers for each question, and the order in which the questions would be asked. Using this, I created a design for the chatbot in Figma. Here I had to be more specific; I designed buttons, color scheme, font, size, etc. Figma also let me see a visual of how the app would work and the transition between pages when buttons were clicked. Once I was satisfied with the design, I used this as a model and started coding on Android Studio. I used the programming language Java to code my app in Android Studio; where I wrote down the questions and values that patients could select from by looking at my first flowchart model. I then designed the app here by importing my designs in Figma to Android Studio. The app took 2 to 3 weeks to code and edit.

The Flow Charts are prepared for all the components and the sample flow charts for Lab Test Results (Table 1) and Frequently Asked Questions (Table 2) are given below.

Table 1: Flow Chart-Lab Test Results.

Table 2: Flow Chart-Frequently Asked



The Coding Sheets were designed. And sample coding sheets (Tables 3-14) are attached in the Appendix I.

Questionnaire Data Analysis using SPSS:

The Statistical analysis namely-Two Sample 't" Test, Chi square test, One-Way ANOVA, Correlation and Factor Analysis are executed for the data derived from the questionnaire.

1. Comparison - DBOT User based on Gender:

The Two Sample 't" Test is done to compare the DBOT user based on Gender and the results are shown in Table 15.

Research hypothesis (H1): There is a significant difference between Gender of the respondent and the user experience of DBOT. Null hypothesis (H0): There is no significant difference between Gender of the respondent and the user experience of DBOT.

Statistical test: Two Sample 't" Test. Level of Significance (α): 0.05

DOI: 10.36838/v4i4.16

Table 15: Two Sample 't' Test, based on Gender.

Components	Male (N = 60)		Female (N =40)		Statistical Inference
	Mean	SD	Mean	SD	
Diabetes History	15.83	3.698	14.91	3.407	T=1.214, Df=98 .228 > 0.05 Not Significant
Medical History	13.29	1.588	13.29	1.545	T=0.020, Df=98 .984 > 0.05 Not Significant
Family History	9.51	1.804	9.34	2.057	T=0.415, Df=98 .679 > 0.05 Not Significant
Medical History (For Women Only)	.00	.000	14.57	1.195	T=-1.722, Df=98 .369 < 0.05 Not Significant
Eating/ Exercising Habits	9.12	.801	9.34	.802	T=-1.308, Df=98 .194 > 0.05 Not Significant
Tracking Health Status (Lab Test Results)	11.45	1.287	11.69	1.157	T=-0.919, Df=98 .361 > 0.05 Not Significant
Medicine Consuming Status	5.43	.558	5.54	.611	T=-0.927, Df=98 .356 > 0.05 Not Significant
Knowledge about reducing the risk	10.37	1.069	10.11	.718	T=1.264, Df=98 .209 > 0.05 Not Significant
Frequently Asked Questions	12.05	3.054	11.97	3.082	T=0.116, Df=98 .908 > 0.05 Not Significant
DBOT Usage	11.06	.768	11.14	.430	T=-0.579, Df=98 .564 > 0.05 Not Significant
All Components	17.46 15	2.58044	16.5429	1.91500	T=1.848, Df=98 .068 > 0.05 Not Significant

Findings:

Table 15 reveals that the Probability value of all components is greater than the level of significant value (0.068 > 0.05), and so the research hypothesis is rejected, and the null hypothesis is accepted. The results prove that there is no significant difference between the DBOT user based on Gender, and hence it could be concluded that the user experience about all Components in the DBOT for Male and Female, does not show a greater difference.

2. Comparison - DBOT User based on Locality:

The Two Sample 't' Test is done to compare the DBOT user based on Locality and the results are shown in Table 16.

Research hypothesis (H1): There is a significant difference between Locality of the respondent and the user experience of DBOT. Null hypothesis (H0): There is no significant difference between Locality of the Respondent and the user experience of DBOT.

Statistical test: Two Sample 't" Test. Level of Significance (α): 0.05

Table 16: Two Sample 't' Test, based on Gender.

Components	City/Town (N = 67)		Village (N =33)		Statistical Inference
	Mean	SD	Mean	SD	
Diabetes History	15.28	3.609	15.97	3.618	T=-0.893, Df=98 .374 > 0.05 Not Significant
Medical History	13.43	1.406	13.00	1.837	T=1.305, Df=98 .195 > 0.05 Not Significant
Family History	9.37	1.976	9.61	1.713	T=-0.578, Df=98 .564 > 0.05 Not Significant
Medical History (For Women Only)	5.36	7.246	4.58	6.615	T=0.522, Df=98 .603 < 0.05 Not Significant
Eating/ Exercising Habits	9.24	.676	9.12	1.023	T=0.686, Df=98 .494 > 0.05 Not Significant
Tracking Health Status (Lab Test Results)	11.61	1.114	11.36	1.475	T=0.939, Df=98 .350 > 0.05 Not Significant
Prescription Status	5.48	.612	5.45	.506	T=0.187, Df=98 .852 > 0.05 Not Significant
Knowledge about reducing the risk	10.28	1.012	10.27	.876	T=0.053, Df=98 .958 > 0.05 Not Significant
Frequently Asked Questions	11.69	2.709	12.70	3.592	T=-1.570, Df=98 .120 > 0.05 Not Significant
DBOT Usage	11.13	.672	11.00	.661	T=0.945, Df=98 .347 > 0.05 Not Significant
All Components	17.1045	2.41921	17.2121	2.39475	T=-0.210, Df=98 .834 > 0.05 Not Significant

Findings:

Table 16 reveals that the Probability value of all components is greater than the level of significant value (0.834 > 0.05) and so the research hypothesis is rejected, and the null hypothesis is accepted. The results prove that there is no significant difference between the DBOT user based on Locality, and hence it could be concluded that the user experience about all components in the DBOT for respondents living in City/Town and Village, does not show a greater difference.

3. Association - DBOT User based on Age:

The Chi-square test is done to check the association between the DBOT user based on Age and the results are shown in Table 17.

Research hypothesis (H1): There is a significant difference between Age of the respondent and the user experience of DBOT. Null hypothesis (H0): There is no significant difference between Age of the Respondent and the user experience of DBOT.

Statistical test: Chi-square test. Level of Significance (α): 0.05

Table 17: Chi-square Test based on Age.

		Age of the Respondents					Statistical Inference
		Diabetes History					
		Below 15	15 - 40	41 - 60	Above 60	Total	
Good	F	10	29	29	2	70	χ^2=81.905 Df=3 .000 < 0.05 Significant
	%	100.0%	96.7%	96.7%	6.7%	70.0%	
Bad	F	0	1	1	28	30	
	%	.0%	3.3%	3.3%	93.3%	30.0%	
		Medical History					
Good	F	0	2	8	11	21	χ^2=11.392 Df=3 .010 < 0.05 Significant
	%	.0%	6.7%	26.7%	36.7%	21.0%	
Bad	F	10	28	22	19	79	
	%	100.0%	93.3%	73.3%	63.3%	79.0%	
		Family History					
Good	F	5	21	20	17	63	χ^2=2.045 Df=3 .563 > 0.05 Not Significant
	%	50.0%	70.0%	66.7%	56.7%	63.0%	
Bad	F	5	9	10	13	37	
	%	50.0%	30.0%	33.3%	43.3%	37.0%	
		Medical History (For Women Only)					
Good	F	5	20	20	20	65	χ^2=1.099 Df=3 .777 > 0.05 Not Significant
	%	50.0%	66.7%	66.7%	66.7%	65.0%	
Bad	F	5	10	10	10	35	
	%	50.0%	33.3%	33.3%	33.3%	35.0%	
		Eating/ Exercising Habits					
Good	F	7	19	23	18	67	χ^2=2.156 Df=3 .541 > 0.05 Not Significant
	%	70.0%	63.3%	76.7%	60.0%	67.0%	
Bad	F	3	11	7	12	33	
	%	30.0%	36.7%	23.3%	40.0%	33.0%	
		Tracking Health Status (Lab Test Results)					
Good	F	10	22	28	23	83	χ^2=7.156 Df=3 .067 > 0.05 Not Significant
	%	100.0%	73.3%	93.3%	76.7%	83.0%	
Bad	F	0	8	2	7	17	
	%	.0%	26.7%	6.7%	23.3%	17.0%	
		Prescription Status					
Good	F	0	8	20	29	57	χ^2=44.920 Df=3 .000 < 0.05 Not Significant
	%	.0%	26.7%	66.7%	96.7%	57.0%	
Bad	F	10	22	10	1	43	
	%	100.0%	73.3%	33.3%	3.3%	43.0%	
		Knowledge about reducing the risk					
Good	F	2	6	5	5	18	χ^2=0.181 Df=3 .981 > 0.05 Not Significant
	%	20.0%	20.0%	16.7%	16.7%	18.0%	
Bad	F	8	24	25	25	82	
	%	80.0%	80.0%	83.3%	83.3%	82.0%	
		Frequently Asked Questions					
Good	F	7	20	17	18	62	χ^2=0.962 Df=3 .810 > 0.05 Not Significant
	%	70.0%	66.7%	56.7%	60.0%	62.0%	
Bad	F	3	10	13	12	38	
	%	30.0%	33.3%	43.3%	40.0%	38.0%	
		DBOT Usage					
Good	F	10	24	23	21	78	χ^2=4.040 Df=3 .257 > 0.05 Not Significant
	%	100.0%	80.0%	76.7%	70.0%	78.0%	
Bad	F	0	6	7	9	22	
	%	.0%	20.0%	23.3%	30.0%	22.0%	
		All Components					
Good	F	9	20	18	14	61	χ^2=6.543 Df=3 .088 > 0.05 Not Significant
	%	90.0%	66.7%	60.0%	46.7%	61.0%	
Bad	F	1	10	12	16	39	
	%	10.0%	33.3%	40.0%	53.3%	39.0%	

Findings:

Table 17 reveals that the Probability value of all components is greater than the level of significant value (0.088 > 0.05) and so the research hypothesis is rejected, and the null hypothesis is accepted. The results prove that there is no significant difference between the DBOT user based on Age, and hence it could be concluded that the user experience about all components in the DBOT for respondents of age groups- <15, 16-40, 41-60, and above 60 years, does not show a greater difference.

Comparison - DBOT User based on Educational Qualification:

The One-Way ANOVA was done to check the comparison between the DBOT user based on Educational Qualification and the results are shown in Table 18.

Research hypothesis (H1): There is a significant difference between Age of the respondent and the user experience of DBOT. Null hypothesis (H0): There is no significant difference between Age of the Respondent and the user experience of DBOT.

Table 18: One-Way ANOVA Test based on Educational Qualification.

Educational Qualification	Mean	SD	SS	DF	MS	Statistical inference
Diabetes History						
Between Groups			34.532	2	17.266	
School (n= 43)	15.30	3.461				F=1.335
College (n = 49)	15.37	3.569				.268 > 0.05
Illiterate (n = 8)	17.50	4.472				Not Significant
Within Groups			1254.458	97	12.933	
Medical History						
Between Groups			1.974	2	.987	
School (n= 43)	13.21	1.552				F=0.398
College (n = 49)	13.29	1.633				.673 > 0.05
Illiterate (n = 8)	13.75	1.282				Not Significant
Within Groups			240.616	97	2.481	
Family History						
Between Groups			3.859	2	1.929	
School (n= 43)	9.58	2.038				F=0.536
College (n = 49)	9.27	1.800				.587 > 0.05
Illiterate (n = 8)	9.88	1.642				Not Significant
Within Groups			348.891	97	3.597	
Medical History (For Women Only)						
Between Groups			54.173	2	27.087	
School (n= 43)	5.12	7.122				F=0.545
College (n = 49)	4.69	6.835				.582 > 0.05
Illiterate (n = 8)	7.50	8.036				Not Significant
Within Groups			4824.827	97	49.740	
Eating/ Exercising Habits						
Between Groups			.980	2	.490	
School (n= 43)	9.26	.658				F=0.754
College (n = 49)	9.20	.912				.473 > 0.05
Illiterate (n = 8)	8.88	.835				Not Significant
Within Groups			63.020	97	.650	
Tracking Health Status (Lab Test Results)						
Between Groups			2.305	2	1.153	
School (n= 43)	11.56	1.333				F=0.742
College (n = 49)	11.43	1.173				.479 > 0.05
Illiterate (n = 8)	12.00	1.195				Not Significant
Within Groups			150.605	97	1.553	
Prescription Status						
Between Groups			1.046	2	.523	
School (n= 43)	5.51	.592				F=1.592
College (n = 49)	5.49	.582				.209 > 0.05
Illiterate (n = 8)	5.12	.354				Not Significant
Within Groups			31.864	97	.328	
Knowledge about reducing the risk						
Between Groups			.873	2	.436	
School (n= 43)	10.19	1.006				F=0.464
College (n = 49)	10.33	.944				.630 > 0.05
Illiterate (n = 8)	10.50	.926				Not Significant
Within Groups			91.287	97	.941	
Frequently Asked Questions						
Between Groups			3.478	2	1.739	
School (n= 43)	11.91	3.365				F=0.184
College (n = 49)	12.02	2.825				.832 > 0.05
Illiterate (n = 8)	12.62	2.875				Not Significant
Within Groups			916.482	97	9.448	
DBOT Usage						
Between Groups			.917	2	.458	
School (n= 43)	11.12	.448				F=1.028
College (n = 49)	11.02	.777				.362 > 0.05
Illiterate (n = 8)	11.38	.916				Not Significant
Within Groups			43.273	97	.446	
All Components						
Between Groups			9.249	2	4.625	
School (n= 43)	16.9302	2.50138				F=0.800
College (n = 49)	17.4286	2.38921				.452 > 0.05
Illiterate (n = 8)	16.5000	1.85164				Not Significant
Within Groups			560.791	97	5.781	

Statistical test: One-Way ANOVA test. Level of Significance (α): 0.05

Findings:

Table 18 reveals that the Probability value of all components is greater than the level of significant value (0.452 > 0.05) and so the research hypothesis is rejected, and the null hypothesis is accepted. The results prove that there is no significant difference between the DBOT user based on Educational Qualification, and hence it could be concluded that the user experience about all components in the DBOT for respondents of Educational Qualification- School completed, College completed and Illiterates, does not show a greater difference.

4. Relationship - DBOT User based on All Components:

The Correlation technique is done to check the relation between the DBOT user based on All Components and the results are shown in Table 19.

Table 19: Correlation based on All Components.

All Components	Diabetes History	Medical History	Family History	Medical History (For Women Only)	Eating/ Exercising Habits	Tracking Health Status (Lab Test Results)	Prescription Status	Knowledge about reducing the risk	Frequently Asked Questions	DBOT Usage
Diabetes History	1	.054	-.018	-.126	-.175	.117	-.495**	.037	.073	.123
		.005	**.008**	**.021**	**.002**	**.024**	**.000**	**.015**	**.046**	**.022**
	100	100	100	100	100	100	100	100	100	100
Medical History	.054	1	.010	.009	-.063	-.002	-.018	-.054	-.035	.110
	.005		**.021**	**.027**	**.036**	**.05**	**.007**	**.05**	**.029**	**.027**
	100	100	100	100	100	100	100	100	100	100
Family History	-.018	.010	1	-.037	-.093	-.129	.045	.019	-.049	-.121
	.008	**.021**		**.015**	**.035**	**.020**	**.05**	**.05**	**.028**	**.032**
	100	100	100	100	100	100	100	100	100	100
Medical History (For Women Only)	-.126	.009	-.037	1	.131	.092	.121	-.112	-.036	.061
	.021	**.027**	**.015**		**.019**	**.036**	**.032**	**.026**	**.019**	**.050**
	100	100	100	100	100	100	100	100	100	100
Eating/ Exercising Habits	-.175	-.063	-.093	.131	1	-.148	.100	.187	-.105	-.128
	.002	**.036**	**.035**	**.019**		**.014**	**.021**	**.002**	**.030**	**.020**
	100	100	100	100	100	100	100	100	100	100
Tracking Health Status (Lab Test Results)	.117	-.002	-.129	.092	-.148	1	-.027	-.226*	.112	.185
	.024	**.05**	**.020**	**.036**	**.014**		**.000**	**.024**	**.026**	**.045**
	100	100	100	100	100	100	100	100	100	100
Prescription Status	-.495**	-.018	.045	.121	.100	-.027	1	-.003	-.086	.099
	.000	**.007**	**.05**	**.032**	**.021**	**.000**		**.007**	**.039**	**.028**
	100	100	100	100	100	100	100	100	100	100
Knowledge about reducing the risk	.037	-.054	.019	-.112	.187	-.226*	-.003	1	.063	.023
	.015	**.05**	**.05**	**.026**	**.002**	**.024**	**.007**		**.031**	**.019**
	100	100	100	100	100	100	100	100	100	100
Frequently Asked Questions	.073	-.035	-.049	-.036	-.105	.112	-.086	.063	1	.158
	.046	**.029**	**.028**	**.019**	**.030**	**.026**	**.039**	**.031**		**.017**
	100	100	100	100	100	100	100	100	100	100
DBOT Usage	.123	.110	-.121	.061	-.128	.185	.099	.023	.158	1
	.022	**.027**	**.032**	**.050**	**.020**	**.045**	**.028**	**.019**	**.017**	
	100	100	100	100	100	100	100	100	100	100

Findings:

Table 19 shows that there is a significant correlation between all the Components namely Diabetes History, Medical History, Family History, Medical History (For Women Only), Eating/ Exercising Habits, Tracking Health Status (Lab Test Results), Prescription Status, Knowledge about reducing the risk, Frequently Asked Questions, and DBOT Usage. From the above analysis, it could be concluded that there exists a

DOI: 10.36838/v4i4.16

strong relationship between all the components among each other.

5. Factor Analysis - DBOT User based on All Components:

The Factor Analysis is done to check the relation between the DBOT user, based on All Components and the results are shown in Tables 20-22.

Table 20: KMO and Bartlett's Test.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.767
Bartlett's Test of Sphericity	Approx. Chi-Square	84.412
	Df	36
	Significance	.000

From the above analysis (Table 20), the KMO value is 0.767, (which is very high), shows that the factor analysis is beneficial to test the components. Bartlett's test results .000 < 0.05, shows that it is significant for each component.

Table 21: Total Variance Explained.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	%of Variance	Cumulative %	Total	%of Variance	Cumulative %	Total	%of Variance	Cumulative %
1	1.784	19.818	19.818	1.784	19.818	19.818	1.758	19.531	19.531
2	1.250	13.892	33.709	1.250	13.892	33.709	1.211	13.456	32.987
3	1.177	13.074	46.783	1.177	13.074	46.783	1.184	13.158	46.145
4	1.098	12.197	58.980	1.098	12.197	58.980	1.104	12.270	58.414
5	1.067	12.022	62.520	1.047	11.631	70.612	1.098	12.197	70.612
6	1.047	11.631	70.612						
7	.957	10.637	81.249						
8	.761	8.459	89.708						
9	.628	6.974	96.682						
10	.299	3.318	100.00						

From the above analysis (Table 21), each component is grouped under 5 factors, according to their flexibility of usage. The first factor has 19.531 percentage of variance, the second factor shows 13.456 percentage of variance, the third factor with 13.158 percentage of variance, the fourth factor having 12.270 percentage of variance and the fifth factor shows 12.197 percentage of variance.

Table 22: Rotated Component Matrix.

Components	Values	Component
Diabetes History	.882	1
Medical History	.889	1
Family History	.642	5
Medical History (For Women Only)	.779	2
Eating/ Exercising Habits	.785	5
Tracking Health Status (Lab Test Results)	.767	2
Prescription Status	.902	4
Knowledge about Reducing the Risk	-.656	3
Frequently Asked Questions	.797	3
DBOT Usage	.889	1

From the above Table 22, it could be derived that the First factor consists of the components- Diabetes History, Medical History and DBOT Usage and the Second factor includes Medical History (For Women Only) and Tracking Health Status (Lab Test Results). The Third factor comprises components- Knowledge about Reducing the Risk and Frequently Asked Questions, the Fourth factor incorporates the component- Prescription Status and the Fifth factor involves components like- Family History and Eating/ Exercising Habits.

The principal findings of this research paper are as follows. The components-Diabetes History, Medical History and DBOT Usage and Very Easy to use by the respondents. The components-Medical History (For Women Only) and Tracking Health Status (Lab Test Results) are Easy to handle and use by the respondents. The components- Knowledge about reducing the risk and Frequently Asked Questions are Moderate to use. The components- Prescription Status is Hard to use and the components-Family History and Eating/ Exercising Habits are Very Hard to use, by the respondents.

Suggestions from the respondents, for contouring DBOT, a User Friendly chatbot:

From the questionnaire, the following proposals and recommendations are delineated by the respondents, to append in DBOT, to make it as a user friendly chatbot.

- Chatbot Content to be displayed in Local language (Tamil- Official language of Tamil Nadu)
- Add a Diet Chart
- Reducing the number of questions under each component
- Increasing the number of questions under each component
- Simplify the FAQ, as some answers are found hard to understand.
- Avoid personal questions
- Recommend Test Centers
- Add Case studies for diabetic complications
- DBOT is extensive and long-drawn
- Encompass Voice Chat Assistance
- Include the Availability of the physician
- Comprehend illustrations and demonstrations for Diabetic Foot Ulcer
- Enclose latest statistics and instructions about Diabetes
- Mapping the close by location of pharmacy
- Set down an Exercise Chart
- Glossary for Diabetes-terms and meaning
- Space to post patient's queries
- Few Monotonous/irrelevant questions to be expunged
- Save option for patients' information
- Include Components relating emotional health
- Network bandwidth is a major bottleneck to access the chatbot
- Include doctor's appointment schedule
- 1 to 1 Conversation with the physician
- Insert "Drag and Drop" option
- Encompass space for Patient's feedback
- Reminder for Follow-up date
- Append Visuals, Smiley's and GIFs
- Avoid Message chunking
- Evade Conversation delays
- Strengthening security for Personal data

■ Conclusion

In this research paper, a Chatbot for Diabetes Mellitus, named DBOT, was designed and shared with the diabetic patients for their utilization. Later, a questionnaire survey was conducted to receive the feedback from the respondents about user experience. I enquired the respondents about what they felt while chatting with DBOT and whether they found anything unclear. The filled-in questionnaire was collected, and the data are analyzed using statistical techniques. The principal findings of this research paper were as follows. According to the results, the components namely-Diabetes History, Medical

History, and DBOT Usage are found to be Very Easy to use by the respondents. The components-Medical History (For Women Only) and Tracking Health Status (Lab Test Results) are Easy to use by the respondents. The components- Knowledge about reducing the risk and Frequently Asked Questions are Moderate in difficulty to use. The components- Prescription Status is hard to use and the components-Family History and Eating/ Exercising Habits are Very Hard to use, by the respondents. Through the questionnaire, proposals and recommendations are delineated by the respondents, to improve in DBOT, to make it as a user friendly chatbot. I was surprised while I gathered the feedback, and I realized those facts that went unnoticed during the building phase. The resulting findings guided me further to structure the DBOT more effectively, in the future. My chatbot does not provide the user with the cure for their diabetes problems but gives patient's data to the doctor. DBOT is a tool supporting the diabetic patients and ensures it will not substitute the professional medical advice that a physician gives in person. It will encourage, stimulate, enlighten, instruct, and captivate the diabetic patients in supervising their health with a simple, manageable, uncomplicated cost-effective technology. This DBOT will provide diabetic patients with clinically validated particulars about Diabetes and all that the patients have to do is to commence chatting with DBOT, being quicker, cheaper, and easier and accessible at all times through a few button clicks. And finally, it connects the user to the doctor if necessary. The future step is to incorporate the recommendations and suggestions of the users and restructure the DBOT for enhanced effectiveness.

Acknowledgements

Harnishya Palanichamy thanks Dr. Rabih Younes, Assistant Professor of the Practice, Department of Electrical and Computer Engineering, Duke University for his continued guidance throughout this project.

References

1. Maria Manuela Cruz-Cunha., Isabel Maria Miranda., & Patricia Gonçalves. (2013, April). Handbook of Research on ICTs and Management Systems for Improving Efficiency in Healthcare and Social Care. https://www.igi-global.com/dictionary/did-health-economics-appear-were/12893.
2. Hartz, J., Yingling, L., & Powell-Wiley, T.M. (2016, December). Use of Mobile Health Technology in the Prevention and Management of Diabetes Mellitus. Current Cardiology. https://pubmed.ncbi.nlm.nih.gov/27826901/.
3. Lamprinos, I., Demski, H., Mantwill, S., Kabak, Y., Hildebrand, C., & Ploessnig, M. (2016, July). Modular ICT-based Patient Empowerment Framework for Self-management of Diabetes: Design perspectives and validation results. International Journal of Medical Information. https://pubmed.ncbi.nlm.nih.gov/27185507/.
4. Fijacko, N., Brzan, P.P., & Stiglic, G. (2015, October). Mobile Applications for Type 2 Diabetes Risk Estimation: A Systematic Review. Journal of Medical System. https://pubmed.ncbi.nlm.nih.gov/26303152/.
5. Szydlo, T., & Konieczny, M. (2016, October). Mobile and Wearable Devices in an Open and Universal System for Remote Patient Monitoring. Microprocessors Microsystem. https://dl.acm.org/doi/abs/10.5555/3034193.3034322f
6. Gao, C., Zhou, L., Liu, Z., Wang, H., & Bowers, B. (2017, May). Mobile Application For Diabetes Self-Management In China: Do they Fit for Older Adults International Journal of Medical Information. https://pubmed.ncbi.nlm.nih.gov/28347449/.
7. Fonda, S., Kedziora, R., Vigersky, R., & Bursell, S. (2010, October) Evolution of a Web-based, Prototype Personal Health Application for Diabetes Self-management. Journal of Bio Medical Information. https://pubmed.ncbi.nlm.nih.gov/20937479/.
8. Hood, M., Wilson, R., Corsica, J., Bradley, L., Chirinos, D., & Vivo, A. (2016, December). What do we know about Mobile Applications for Diabetes Self-management? Journal of Behavior Medicine. https://pubmed.ncbi.nlm.nih.gov/27412774/.
9. American Diabetes Association. Frequently Asked Questions: COVID-19 and Diabetes. https://www.diabetes.org/coronavirus-covid-19/how-coronavirus-impacts-people-with-diabetes.
10. Vijayaprasad Gopichandran & Kalirajan Sakthivel. (2021, June). PLOS ONE. Doctor-Patient Communication and Trust in Doctors during COVID-19 times-A Cross Sectional Study in Chennai, India. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0253497.
11. Mohan, V. (2020, December). Health World.com. The Risk of a Fatal Outcome from COVID-19 is up to 50% Higher in People with Diabetes. https://economictimes.indiatimes.com/industry/miscellaneous/risk-of-fatal-outcome-from-covid-19-is-up-to-50-percent-higher-in-people-with-diabetes/diabetics-at-risk/slideshow/79591019.cms.
12. Sarah, L., & Medha N. (2020, July). JAMA Internal Medicine. Caring for Older Adults with Diabetes during the COVID-19 Pandemic.https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2768362.
13. Pawar, P., Jones, V., Bert.-Jan F. van Beijnum, & Hermens, H. (2012, June). A Framework for the Comparison of Mobile Patient Monitoring Systems. Journal of Bio Medical Information. https://www.sciencedirect.com/science/article/pii/S1532046412000287.
14. Farahani, B., Firouzi,F., Chang, V., Badaroglu, M., Constant, N., & Mankodiya, K. (2018, January). Towards Fog-driven IoT eHealth: Promises and Challenges of IoT in Medicine and Healthcare. Future Generation Computer System. https://www.sciencedirect.com/science/article/abs/pii/S0167739X17307677.
15. Andrew Reyner Wibowo Tjiptomongsoguno., Audrey Chen., Hubert Michael Sanyoto., Edy Irwansyah., & Bayu Kanigoro. Medical Chatbot Techniques: A Review, Software Engineering Perspectives in Intelligent Systems, Book series AISC, volume 1294, Springer International Publishing, 2020, November, pp 1-11. DOI: 10.1007/978-3-030-63322-6_28.
16. Vivek Katariya., & Vitthal, S. Intelligent Healthbot for Transforming Healthcare. Proceedings of National Conference on Machine Learning, Department of Information Technology, MIT College of Engineering, Pune, India, March 26, 2019. ISBN: 978-93-5351-521-8. (Accessed on 29 October 2021). https://www.researchgate.net/publication/332413616_Intelligent_Healthbot_for_Transforming_Healthcare
17. Ahmed Fadhil. (2018, March). A Conversational Interface to Improve Medication Adherence: Towards AI Support in Patient's Treatment. https://www.researchgate.net/publication/324055736_A_Conversational_Interface_to_Improve_Medication_Adherence_Towards_AI_Support_in_Patient's_Treatment.
18. Engati Team. How are Intelligent Healthcare Chatbots being used in 2021 and Beyond? (2021 May). https://www.engati.com/blog/chatbots-for-healthcare.
19. Nadarzynski, T., Miles, O., Cowie, A., & Ridge, D. (2019 August). Acceptability of Artificial Intelligence (AI)-led Chatbot Services in Healthcare: A Mixed-methods Study. Digital Health.

DOI: 10.36838/v4i4.16

https://pubmed.ncbi.nlm.nih.gov/31467682/.
20.Dharwadkar, R., & Deshpande, N.A. (2018 June). A Medical Chatbot. International Journal of Computer Trends Technology. https://www.ijcttjournal.org/archives/ijctt-v60p106.
21.Krishnendu Rarhi., Abhishek Mishra., & Krishnasis Mandal. (2017 January). Automated Medical Chatbot. SSRN Electronic Journal. https://www.researchgate.net/publication/326469944_Automated_Medical_Chatbot.
22.Palanica, A., Flaschner, P., Thommandram, A., Li, M., & Fossat, Y. (2019 April). Physicians' Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey. Journal of Medical In ternet Research. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6473203/.
23.Shaker El-Sappagh Farman Ali., Samir El-Masri., Kyehyun Kim., Amjad Ali., & Kyung-Sup Kwak. (2018 November). Mobile Health Technologies for Diabetes Mellitus: Current State and Fu ture Challenges. IEEE Access. https://www.researchgate.net/publication/328951990_Mobile_Health_Technologies_for_Diabetes_Mellitus_Current_State_and_Future_Challenges.
24.Abbas Saliimi Lokman., & Jasni Mohamad Zain. (2007 August). Designing a Chatbot for Diabetic Patients. https://www.researchgate.net/publication/266872926_Designing_a_Chatbot_for_Diabetic_Patients.
25.Aishwarya Kedar., Jyoti Dahale, Khushboo Patel., Shivani La hamage., Chordiya., S.G. (2020 September). Chatbot System for Healthcare using Artificial Intelligence. International Journal of Scientific Development and Research. https://www.ijsdr.org/papers/IJSDR2009083.pdf.
26.Bernhaupt., R. Evaluating User Experience in Games: Concepts and Methods. Springer, London, 2010. https://link.springer.com/book/10.1007%2F978-1-84882-963-3.
27.Asbjon Folstad., & Petter Bae Brandtzaeg. (2019 May). User Ex periences with Chatbots: Findings from a Questionnaire Study. Quality and User Experience. https://doi.org/10.1007/s41233-020-00033-2.
28.Mariska E Te Pas., Werner G M M Rutten., Arthur Bouwman, R., & Marc P Buise. (2020 December). User Experience of a Chat bot Questionnaire Versus a Regular Computer Questionnaire: Prospective Comparative Study. JMIR Med Inform. https://pubmed.ncbi.nlm.nih.gov/33284125/.
29.Zeljko Maric. (2018 March). The User Experience of Chatbots. A Design Science Approach. https://research-api.cbs.dk/ws/portalfiles/portal/59778977/499371_Master_Thesis_Zeljko_Maric.pdf.
30.Nogueras.A. (2021). Evaluating Chatbot User Experience-Digital WPI. https://digital.wpi.edu/downloads/t722hc79t.

Author

Harnishya Palanichamy is in Grade 10 at Hebron School, Ootacamund, Tamil Nadu, India. She's passionate about the field of computer science and is fluent in the coding languages; Java and Java script. She enjoys coding and researching about AI. She started her coding journey by coding games in Java script, and she also has experience with robotics; being in the school robotics club. In the future, she wants to develop her coding knowledge by creating more complex apps.

Appendix I

Table 3: AndriodManifest

AndriodManifest.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<manifest package="com.example.chatbotexample"
xmlns:android="http://schemas.android.com/apk/res/android"><application
android:theme="@style/Theme.ChatbotExample" android:supportsRtl="true"
android:roundIcon="@mipmap/ic_launcher_round" android:label="@string/app_name"
android:icon="@mipmap/ic_launcher" android:allowBackup="true"><activity
android:name=".homepage"/><activity android:name=".MainActivity"><intent-
filter><action android:name="android.intent.action.MAIN"/><category
android:name="android.intent.category.LAUNCHER"/></intent-
filter></activity></application></manifest>
```

Table 4: Activity_main.

Activity_main

```
<?xml version="1.0" encoding="UTF-8"?>
<androidx.constraintlayout.widget.ConstraintLayout tools:context=".MainActivity"
android:layout_height="match_parent" android:layout_width="match_parent"
xmlns:tools="http://schemas.android.com/tools"
xmlns:app="http://schemas.android.com/apk/res-auto"
xmlns:android="http://schemas.android.com/apk/res/android"><ImageView
android:layout_height="150dp" android:layout_width="150dp"
app:srcCompat="@drawable/ic_launcher" app:layout_constraintVertical_bias="0.521"
app:layout_constraintTop_toTopOf="parent"
app:layout_constraintStart_toStartOf="parent"
app:layout_constraintHorizontal_bias="0.563" app:layout_constraintEnd_toEndOf="parent"
app:layout_constraintBottom_toBottomOf="parent"
android:id="@+id/imageView"/><TextView android:layout_height="wrap_content"
android:layout_width="wrap_content" app:layout_constraintVertical_bias="0.636"
app:layout_constraintTop_toTopOf="parent" app:layout_constraintHorizontal_bias="0.496"
app:layout_constraintBottom_toBottomOf="parent" android:id="@+id/textView"
app:layout_constraintRight_toRightOf="parent"
app:layout_constraintLeft_toLeftOf="parent" android:text="Get your questions answered
with DBot"/><Button android:layout_height="wrap_content"
android:layout_width="wrap_content" app:layout_constraintVertical_bias="0.71"
app:layout_constraintTop_toTopOf="parent"
app:layout_constraintStart_toStartOf="parent"
app:layout_constraintEnd_toEndOf="parent"
app:layout_constraintBottom_toBottomOf="parent" android:id="@+id/button"
android:text="Get Started" app:backgroundTint="#3DBFBB"
android:textColorHighlight="#8F1B1B" android:textColor="#F3ECEC"
android:textAllCaps="false"/><ImageView android:layout_height="319dp"
android:layout_width="531dp" app:srcCompat="@drawable/background"
app:layout_constraintVertical_bias="0.0" app:layout_constraintTop_toTopOf="parent"
app:layout_constraintStart_toStartOf="parent"
app:layout_constraintEnd_toEndOf="parent"
app:layout_constraintBottom_toBottomOf="parent" android:id="@+id/imageView2"
android:layout_margin="-10dp"/></androidx.constraintlayout.widget.ConstraintLayout>
```

Table 5: Activity_homepage.

Activity_homepage

```
<?xml version="1.0" encoding="UTF-8"?>
<androidx.constraintlayout.widget.ConstraintLayout tools:context=".homepage"
android:layout_height="match_parent" android:layout_width="match_parent"
xmlns:tools="http://schemas.android.com/tools"
xmlns:app="http://schemas.android.com/apk/res-auto"
xmlns:android="http://schemas.android.com/apk/res/android"><androidx.recyclerview.widget.RecyclerView android:layout_height="match_parent"
android:layout_width="match_parent" tools:listitem="@layout/textin"
android:background="@color/black" android:id="@+id/recycler_view"/> //edit text and
button </androidx.constraintlayout.widget.ConstraintLayout>
```

Table 6: Activity_labtestchat.

Activity_labtestchat

```
<?xml version="1.0" encoding="UTF-8"?>
<androidx.constraintlayout.widget.ConstraintLayout
xmlns:app="http://schemas.android.com/apk/res-auto"
xmlns:tools="http://schemas.android.com/tools" android:background="@color/white"
android:layout_height="match_parent" android:layout_width="match_parent"
xmlns:android="http://schemas.android.com/apk/res/android"><LinearLayout
android:layout_height="match_parent" android:layout_width="match_parent"
app:layout_constraintTop_toTopOf="parent"
app:layout_constraintRight_toRightOf="parent"
app:layout_constraintLeft_toLeftOf="parent"
app:layout_constraintBottom_toTopOf="@id/bottomlayout"
android:id="@+id/linearcompact"><androidx.recyclerview.widget.RecyclerView
android:layout_height="wrap_content" android:layout_width="match_parent"
android:id="@+id/recycler_viewLabTest"
android:layout_marginBottom="100dp"/></LinearLayout><LinearLayout
android:layout_height="wrap_content" android:layout_width="match_parent"
android:id="@+id/bottomlayout" android:layout_weight="0" android:visibility="visible"
android:weightSum="2" app:layout_constraintBottom_toBottomOf="@+id/linearcompact"
android:padding="10dp"><com.google.android.material.textfield.TextInputLayout
android:layout_height="wrap_content" android:layout_width="match_parent"
app:layout_constraintTop_toTopOf="parent" android:id="@+id/ChatextField"
android:layout_weight="2" app:layout_constraintBottom_toBottomOf="parent"
app:layout_constraintVertical_bias="0.269" app:layout_constraintStart_toStartOf="parent"
app:layout_constraintHorizontal_bias="0.535" app:layout_constraintEnd_toEndOf="parent"
android:layout_marginRight="10dp" android:hint="Enter Your Message Here"
style="@style/Widget.MaterialComponents.TextInputLayout.OutlinedBox"><com.google.android.material.textfield.TextInputEditText android:layout_height="match_parent"
android:layout_width="match_parent"
android:id="@+id/chatcontentforsending"/></com.google.android.material.textfield.TextInputLayout><com.google.android.material.floatingactionbutton.FloatingActionButton
android:background="@color/chat_view_incoming"
android:layout_height="match_parent" android:layout_width="match_parent"
android:id="@+id/button_submit" app:layout_constraintBottom_toBottomOf="parent"
app:layout_constraintStart_toStartOf="parent"
app:layout_constraintEnd_toEndOf="parent" app:tint="@color/white"
app:backgroundTint="@color/chat_view_incoming"
android:src="@drawable/ic_baseline_send_24"
android:layout_margin="2dp"/></LinearLayout></androidx.constraintlayout.widget.ConstraintLayout>
```

Table 5: Activity_faqpage.

Activity_faqpage

```
<?xml version="1.0" encoding="UTF-8"?>
<androidx.constraintlayout.widget.ConstraintLayout
xmlns:tools="http://schemas.android.com/tools" android:background="@color/white"
android:layout_height="match_parent" android:layout_width="match_parent"
xmlns:android="http://schemas.android.com/apk/res/android"><androidx.recyclerview.wi
dget.RecyclerView android:layout_height="match_parent"
android:layout_width="match_parent"
android:id="@+id/recycler_viewFAQ"/></androidx.constraintlayout.widget.ConstraintLayo
ut>
```

Table 8: testin.

Textin

```
<?xml version="1.0" encoding="UTF-8"?>
<RelativeLayout android:id="@+id/main_layout" android:layout_height="wrap_content"
android:layout_width="wrap_content" xmlns:tools="http://schemas.android.com/tools"
xmlns:android="http://schemas.android.com/apk/res/android"><RelativeLayout
android:id="@+id/message_layout" android:layout_height="wrap_content"
android:layout_width="wrap_content" android:background="@android:color/white"
android:padding="10dp" android:layout_margin="8dp"><TextView
android:id="@+id/message_text" android:layout_height="wrap_content"
android:layout_width="wrap_content" android:text="Hi"
android:textColor="@android:color/black" android:textSize="16sp"
android:layout_alignParentLeft="true" android:maxWidth="250dp"
android:minWidth="120dp"/><TextView android:id="@+id/date_text"
android:layout_height="wrap_content" android:layout_width="wrap_content"
tools:text="11:34" android:textColor="#b3b3b3" android:textSize="11sp"
android:visibility="gone" android:layout_below="@+id/message_text"
android:layout_alignRight="@+id/message_text"/><LinearLayout
android:id="@+id/VerticalLinearLayout" android:layout_height="wrap_content"
android:layout_width="match_parent" android:visibility="gone"
android:orientation="vertical"><Button android:id="@+id/LabTestsButton"
android:layout_height="wrap_content" android:layout_width="wrap_content"
android:text="Lab Test Results" android:textAllCaps="false"/><Button
android:id="@+id/FAQsButton" android:layout_height="wrap_content"
android:layout_width="wrap_content" android:text="FAQs"
android:textAllCaps="false"/></LinearLayout></RelativeLayout></RelativeLayout>
```

Table 9: testout.

Textout

```
<?xml version="1.0" encoding="UTF-8"?>
<RelativeLayout android:id="@+id/main_layout" android:layout_height="wrap_conten
android:layout_width="wrap_content" xmlns:tools="http://schemas.android.com/tools"
xmlns:android="http://schemas.android.com/apk/res/android"><RelativeLayout
android:id="@+id/message_layout" android:layout_height="wrap_content"
android:layout_width="wrap_content" android:layout_alignParentRight="true"
android:background="#5089fa" android:padding="10dp"
android:layout_margin="8dp"><TextView android:id="@+id/message_text"
android:layout_height="wrap_content" android:layout_width="wrap_content"
android:text="Hi" android:textColor="@android:color/white" android:textSize="16sp"
android:layout_alignParentLeft="true" android:maxWidth="250dp"
android:minWidth="120dp"/><TextView android:id="@+id/date_text"
android:layout_height="wrap_content" android:layout_width="wrap_content"
tools:text="11:34" android:textColor="@android:color/white" android:textSize="11sp"
android:visibility="gone" android:layout_below="@+id/message_text"
android:layout_alignRight="@+id/message_text"/></RelativeLayout></RelativeLayout>
```

Table 10: CustomAdapter.

CustomAdapter.java

```
package com.example.chatbotexample;
import android.content.Context;
import android.view.LayoutInflater;
import android.view.View;
import android.view.ViewGroup;
import android.widget.LinearLayout;
import android.widget.TextView;
import androidx.recyclerview.widget.RecyclerView;
import java.text.DateFormat;
import java.util.ArrayList;
public class CustomAdapter extends RecyclerView.Adapter<RecyclerView.ViewHolder> {
    private final Context context;
    ArrayList<MessageModel> list;
    public static final int MESSAGE_TYPE_IN = 1;
    public static final int MESSAGE_TYPE_OUT = 2;
    public CustomAdapter(Context context, ArrayList<MessageModel> list) { // you can pass other parameters in constructor
        this.context = context;
        this.list = list;
    }
private class MessageInViewHolder extends RecyclerView.ViewHolder {
TextView messageTV,dateTV;
 LinearLayout linearbuttonLayout;
  MessageInViewHolder(final View itemView) {
          super(itemView);
          messageTV = itemView.findViewById(R.id.message_text);
          dateTV = itemView.findViewById(R.id.date_text);
          linearbuttonLayout= itemView.findViewById(R.id.VerticalLinearLayout);
      }
      void bind(int position) {
          MessageModel messageModel = list.get(position);
          if(messageModel.message.equals("LAYOUTCHANGE"))
          {
              linearbuttonLayout.setVisibility(View.VISIBLE);
              messageTV.setVisibility(View.GONE);
          }
          else{
              linearbuttonLayout.setVisibility(View.GONE);
              messageTV.setVisibility(View.VISIBLE);
          }
messageTV.setText(messageModel.message);
          dateTV.setText(DateFormat.getTimeInstance(DateFormat.SHORT).format(messageModel.messageTime));
      }
    }
    private class MessageOutViewHolder extends RecyclerView.ViewHolder {
        TextView messageTV,dateTV;
        MessageOutViewHolder(final View itemView) {
            super(itemView);
            messageTV = itemView.findViewById(R.id.message_text);
            dateTV = itemView.findViewById(R.id.date_text);
        }
        void bind(int position) {
            MessageModel messageModel = list.get(position);
            messageTV.setText(messageModel.message);
            dateTV.setText(DateFormat.getTimeInstance(DateFormat.SHORT).format(messageModel.messageTime));
        }
    }
    @Override
    public RecyclerView.ViewHolder onCreateViewHolder(ViewGroup parent, int viewType) {
        if (viewType == MESSAGE_TYPE_IN) {
            return new MessageInViewHolder(LayoutInflater.from(context).inflate(R.layout.textin, parent, false));
        }
        return new MessageOutViewHolder(LayoutInflater.from(context).inflate(R.layout.textout, parent, false));
    }
    @Override
    public void onBindViewHolder(RecyclerView.ViewHolder holder, int position) {
        if (list.get(position).messageType == MESSAGE_TYPE_IN) {
            ((MessageInViewHolder) holder).bind(position);
        } else {
            ((MessageOutViewHolder) holder).bind(position);
        }
    }
    @Override
    public int getItemCount() {
```

Table 11: Question And Answer Variable.

QuestionAndAnswerVariable.java

```
package com.example.chatbotexample;
public interface QuestionAndAnswerVariable {
   String Question1 = "Welcome to DBot";
   String Question2 = "Choose from the options below to start the chat";
   String QuestionBloodPressureLevel = "Enter your Blood Pressure Level";
   //HomePage Questions
   String QuestionBasicOne= "Welcome App";
   String QuestionBasicTwo = "Enter your Name";
   //BloodTest Selection Area
   String QuestionForChooseTest = "Please choose the test result you would like to check first";
   String AnswerForChooseTest1 = "Blood Pressure Level";
   String AnswerForChooseTest2 = "Blood Oxygen Level";
   String AnswerForChooseTest3 = "Fasting Blood Sugar Level";
```

Table 12: Message Model.

MessageModel.java

```
package com.example.chatbotexample;

import java.util.Date;

public class MessageModel {

    public String message;

    public int messageType;

    public Date messageTime = new Date();

    // Constructor

    public MessageModel(String message, int messageType) {

        this.message = message;

        this.messageType = messageType;

    }

}
```

Table 13: homepage.

Homepage.java

```
package com.example.chatbotexample;

import androidx.appcompat.app.AppCompatActivity;

import androidx.recyclerview.widget.LinearLayoutManager;

import androidx.recyclerview.widget.RecyclerView;

import android.os.Bundle;

import java.util.ArrayList;

public class homepage extends AppCompatActivity {

  RecyclerView recyclerView;

  @Override

  protected void onCreate(Bundle savedInstanceState) {

      super.onCreate(savedInstanceState);

      setContentView(R.layout.activity_homepage);

      ArrayList<MessageModel> messagesList = new ArrayList<>();

      messagesList.add(new
MessageModel(QuestionAndAnswerVariable.Question1,CustomAdapter.MESSAGE_TYPE_IN));

      messagesList.add(new
MessageModel(QuestionAndAnswerVariable.Question2,CustomAdapter.MESSAGE_TYPE_IN));

      messagesList.add(new MessageModel("LAYOUTCHANGE",CustomAdapter.MESSAGE_TYPE_IN));

      messagesList.add(new MessageModel(" Android",CustomAdapter.MESSAGE_TYPE_IN));

      messagesList.add(new MessageModel("Reply Studio",CustomAdapter.MESSAGE_TYPE_OUT));

//      for (int i=0;i<10;i++) {

//          messagesList.add(new MessageModel("Hi", i % 2 == 0 ?
CustomAdapter.MESSAGE_TYPE_IN : CustomAdapter.MESSAGE_TYPE_OUT));

//      }

      CustomAdapter adapter = new CustomAdapter(this, messagesList);

      recyclerView = findViewById(R.id.recycler_view);

      recyclerView.setLayoutManager(new LinearLayoutManager(this));

      recyclerView.setAdapter(adapter);

  }

}
```

Table 14: MainActivity.

MainActivity.java

```
package com.example.chatbotexample;

import androidx.appcompat.app.AppCompatActivity;

import android.content.Intent;

import android.os.Bundle;

import android.view.View;

import android.widget.Button;

import android.widget.TextView;

public class MainActivity extends AppCompatActivity {

  Button getStarted ;

  TextView questionsAnswered ;

  @Override

  protected void onCreate(Bundle savedInstanceState) {

    super.onCreate(savedInstanceState);

    setContentView(R.layout.activity_main);

    getStarted = findViewById(R.id.button);

    questionsAnswered= findViewById(R.id.textView);

    getStarted.setOnClickListener(new View.OnClickListener() {

      @Override

      public void onClick(View view) {

        Intent mainToHome = new Intent(MainActivity.this,homepage.class);

        //mainToHome.

        startActivity(mainToHome);

      }

    });

  }

}
```


DOI: 10.36838/v4i3.16

IJHSR
ijhighschoolresearch.org
■ RESEARCH ARTICLE

Enhancing the Efficiency of a Plant Microbial Fuel Cell through the Use of Methane-Oxidizing Bacteria

SeungHan Ha, Daniel Seungmin Lee
Chadwick International, 45 Art center-daero 97beon-gil, Songdo-dong, Yeonsu-gu, Incheon, 22002, Republic of Korea; shseanha@gmail.com

ABSTRACT: Methane-oxidizing bacteria (MOB) and soil microorganisms (SMOs) are known to decompose methane and root exudates, respectively, to produce electrons and soil nutrients like ammonium. This investigation examines how the combined use of the two species may increase the efficiency of existing plant microbial fuel cells (PMFCs). It was found that power output increased by 10.2% when *Methylomonas sp*. (a type of MOB) was used with SMO A separated from local paddy soil. Additionally, SMO A was shown to enhance the growth of both cabbage and rice plants, while *Methylomonas sp*. had little influence on cabbage. When the two species were individually added to PMFCs, SMO A showed more sustained electrical production compared to that of *Methylomonas sp*. Power output of the PMFC including the MOB increased on days 2 and 3, but rapidly decreased as the oxygen composition of the soil decreased. For this reason, MOB seem to be ineffective as a primary, long-term electron source in PMFCs, though it can increase efficiency by providing nutrients from methane oxidation to *Geobacter*. MOB's combination with *Geobacter* seems promising as it creates an anoxic environment (suitable to *Geobacter*) by consuming oxygen.

KEYWORDS: Electrochemistry; Alternative Energy Sources; Microbial Fuel Cells; Plant Microbial Fuel Cells (PMFCs); Methane-Oxidizing Bacteria.

■ Introduction

With current advancements in technology, the global consumption of electricity is showing unprecedented growth. It was even predicted that electricity usage will increase by over 25% from 2017 to 2040.[1] However, current means of electricity production display several environmental and economic problems. At the current rate of population growth, the development of an efficient, eco-friendly, and economic energy source is of urgent need, and Plant Microbial Fuel Cells (PMFCs) seem to be a promising candidate.[2,3] Solely reliant on plants and microorganisms to decompose nutrients, PMFCs are beneficial in that they require minimal industrial input. The only practical caveat of PMFCs is their low efficiency (typically around 0.022%).[4]

A plant microbial fuel cell system involves multiple steps. First, when soil microorganisms oxidize organic substances from the plant in the PMFC by dissociating them near the cathode, electrons are produced. Then the electrons near the cathode move to the anode and reduce oxygen, an oxidizing agent; current is generated from the movement of electrons. The type of microorganism present in the soil can affect the oxidation reactions and thus the power efficiency of a PMFC. The following equations show examples of electrons being produced from the oxidation of glucose and acetate respectively.[5]

$$C_6H_{12}O_6 + 6H_2O \rightarrow 6CO_2 + 24H^+ + 24e^-$$

OR

$$CH_3COO^- + 3H_2O \rightarrow CO_2 + HCO_3^- + 8H^+ + 8e^-$$

Given the information above, this study examines the effects of introducing a new microorganism – specifically using the synergism between the existing soil microorganisms and methane-oxidizing bacteria – to enhance the power efficiency of the current PMFC model.

All MOB can be categorized into two groups: anaerobic and aerobic (or facultatively anaerobic). Anaerobic MOB have been incorporated in many PMFC-related studies using rice plants in paddy fields because the bottom layer of paddy fields is favorable to anaerobic bacteria and rice farming is a type of hydroponic system suitable for electrons to flow. Yet, aerobic MOB also have the potential to be involved in PMFC research because paddy fields have layers that consist of 10% oxygen. In essence, MOB go through the following oxidation reaction to produce electrons.[6]

$$CH_4 + 3O_2^- \rightarrow CO + 2H_2O + 6e^-$$

OR

$$CH_4 + 4O_2^- \rightarrow CO_2 + 2H_2O + 8e^-$$

In addition to producing electrons, this oxidation reaction produces biofuels, such as methanol, and other by-products, such as CH_2O and $HCOO^-$, that are involved in the oxidation reaction of Geobacter, producing even more electrons.[2] Consequently, the current is increased and thus power efficiency of the PMFC is enhanced, further illustrating the potential of aerobic MOB.

Moreover, certain types of MOB appear to be beneficial to plant growth as a previous study claims that treating peanuts with MOB aided the growth of their roots and root hairs.[7] Improvements in plant growth also can increase the power production of PMFCs as more organic material will be produced by the plant.

The aim of this investigation is to observe the effects of aerobic MOB and SMO on plant growth to develop a more power efficient PMFC model using the two microorganisms. First, the symbiosis of MOB and SMOs in local soil was investigated to determine the appropriate microbe pair and ratio for power output. Next, with the selected MOB and SMO pair, their effects on the growth of rice plants were

DOI: 10.36838/v4i4.17

observed. Lastly, the final model was applied to paddy soil with rice plants.

Methods

MOB Incubation and Methane Gas Production:

Initially, two types of MOB – *Methylosinus trichosporium* (KCTC 12760) and *Methylomonas sp.* (KCTC 62176) – were chosen to be used in this study because they can both survive in aerobic conditions and have been studied in the past. The bacteria culture medium that was used consisted of 2.94 g of ammonium mineral salt (AMS, MB cell MB-A0722) and 5 mL of methanol dissolved in 1 L of distilled water. The solution was autoclaved at 121°C for 15 minutes and was cooled down to 22~25°C before use. Each type of bacteria was streaked on the solid AMS culture medium in a petri dish using an inoculation loop and then was cultured at 30°C in an incubator. Figure 1 shows the resulting bacterial colonies.

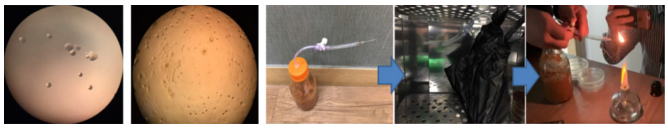


Figure 1: From the left: Incubated *Methylosinus trichosporium* colonies, incubated *Methylomonas sp.* colonies, methane gas synthesis using food waste, checking the presence of methane gas.

A series of preliminary experiments was then performed to design a system for supplying methane gas to the two MOB. The system involved attaching a rubber tube, a small valve, and an injection needle to the lid of a food waste bin and fermenting the waste for seven days. The end of the needle was lit on fire to confirm whether methane was being produced in the bin.

Selection of MOB:

While supplying methane to the MOB, the changes in the voltage of the cell and the growth rate of the rice plants were observed. Both sterilized and unsterilized paddy soil were used to examine the impact of existing soil microorganisms on the voltage of the cell. The electrical cell used in this experiment had a zinc plate as its anode and a copper plate as its cathode, as shown in Figure 2. To supply a consistent amount of methane, the food waste bin was connected to each cell using plastic pipes attached to one-way valves. Each cell contained the AMS culture medium, MOB 1 or 2, and a sponge with 5 rice seeds.

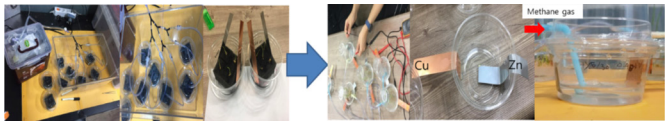


Figure 2: (Left) evenly supplying methane gas from the food waste bin to each container. (Right) Measuring voltage across each MOB container after supplying it with methane gas.

In the preliminary experiments, it was observed that fewer and shorter rice seeds sprouted in the cells containing MOB 1 than in those with MOB 2, as shown in Figure 3. The average voltage was also lower in the cells containing MOB 1 than in the control groups or those with MOB 2. Moreover, the average voltage was higher in the cells containing soil microorganisms and was increased even further when MOB was added. Consequently, the soil microorganisms and MOB 2 were used in the subsequent experiments, and MOB 1 was eliminated.

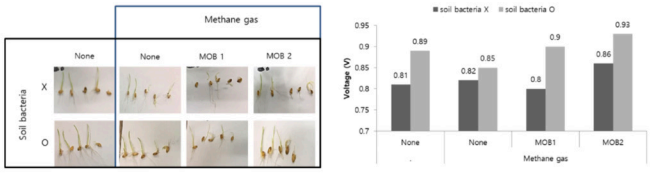


Figure 3: Different combinations of SMO and MOB species and resulting effects on rice plant growth.

Classifying SMOs Based on Electrochemical Characteristics:

The soil microorganism most favorable to the experiment was found by isolating each type of bacteria in the paddy soil that was brought from South Jeolla Province and examining the effect of each type of bacteria on the voltage of the cell and growth of rice plants. 1 g of paddy soil and 10 mL of distilled water were mixed and streaked with an inoculation loop on a solid NB culture medium. The bacteria were grouped into five types based on appearance as shown in Figure 4.

The five types of soil bacteria were separately cultured again

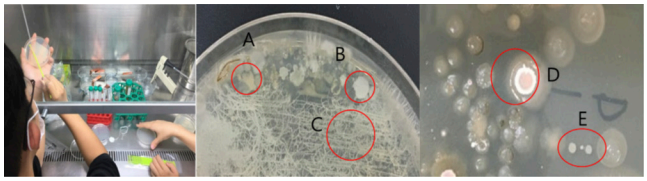


Figure 4: Incubating and separating soil microorganism colonies into 5 groups (A~E).

on the NB medium. On Day 0 and Day 3 of culturing, the absorbance of the bacteria and the voltage of the cell were observed using a spectrophotometer and multimeter, respectively.

Electrochemical Behavior of SMOs when Incubated with MOB:

The initial concentrations of SMO A~E were set to be equal at absorbance 0.1 AU before being cultured for 3 days. The NB medium with the cultured SMOs was placed in a 30°C water bath. A salt bridge was created by heating a 2 M KNO3 solution with agar to measure electrical properties of the solution. A 100 Ω resistor was connected in series to the SMO electrochemical cell to measure power output for 5 minutes each. Circuit diagrams are shown in Figure 5.

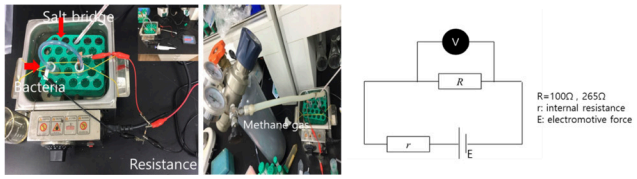


Figure 5: Measuring power output of each cell containing SMO and MOB.

It was found that power output was the highest for the cell containing SMO A. Increase in power of SMO A in a NB medium was measured. As the addition of microorganisms could have influenced the electrical properties (i.e., conductivity) of the solution, its impact was monitored by measuring power output directly from the terminals of the cells without the resistor. Finally, methane gas was readily injected into a solution containing NB, AMS, and MOB (Methylomonas sp.) was cultured, as cultured SMO A was periodically introduced.

DOI: 10.36838/v4i4.17

The addition of SMO A in the solution was shown to increase total power output.

Symbiosis of SMO and MOB Based on Effects on Plant Growth and Electrochemical Behavior:

To determine the optimal ratio and concentration of the solution containing soil bacteria A and MOB, three ten-fold serial dilutions (1:10, 1:100, 1:1000) were performed. Then each diluted solution was treated onto a solid 0.7% agar medium in a petri dish, and 16 cabbage seeds were planted in each dish. A week later, the chlorophyll concentration of the cabbage plants and the voltage of each petri dish containing the different solutions were measured.

Application of SMO+MOB Microbial Fuel Cells to Rice Plants in Paddy Soil:

Rice plant seeds were germinated in 0.7% agar medium in advance and replanted in paddy soil. A week after treating SMO A and MOB in each plastic container with paddy soil as shown in Figure 6, the length of the rice plants in and the voltage of each container were measured. However, there was no significant change in the voltage, so the voltage of each container was measured again but on a daily basis for a week. Starting with this experiment, a pure methane gas container was used instead of the food waste bin to supply methane to the fuel cell.

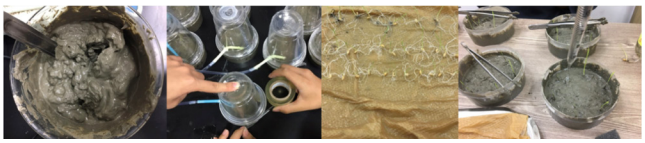


Figure 6: Preparing rice paddy and rice plant containers with SMO A and MOB (*Methylomonas sp.*).

The pH level of the paddy soil in each container was measured using a pH meter and once again using phenol red. The phenol red test was performed by measuring the absorbance of the paddy soil in each container at a wavelength of 415 nm using a spectrophotometer.

Results and Discussion

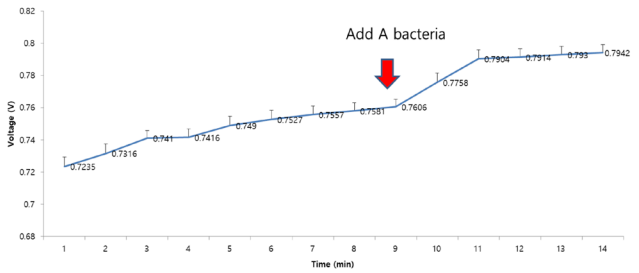


Figure 7: Electrochemical behavior of SMOs when incubated with MOB: potential difference across the 100Ω resistor by time.

Before the addition of MOB and SMO A, voltage remained at around 0.741 V. With the addition of MOB then SMO A, voltage slightly increased to 0.750 V on average then to around 0.794 V, as shown in Figure 7. The addition of MOB caused a 7.0×10^{-5} mW increase in power and 1.1×10^{-5} mW when SMO A was added to it. It was noted that the voltage reading dropped when bacteria were added. These changes were likely due to surface disturbances as the bacteria were added. Such effects were removed from the data as outliers.

The results of the experiment support the initial hypothesis that the presence of both SMO and MOB increases electrical output. However, with further considerations, the reasoning seemed misleading. The voltage of the cell must be constant regardless of biological activity as there is a set potential difference for every cathode-anode pair. Hence, it was that, with more free electrons in the solution due to the oxidation, the conductivity of the solution increased, decreasing the cell's internal resistance. After the experiment, it seemed likely that the increase in voltage measurements of the cell was a result of this decrease in internal resistance: $V=\varepsilon-IR_{internal}$. This new explanation is valid in the sense that less energy (potential drop) is dissipated by the internal resistance, and power output increases when SMO and MOB are used together.

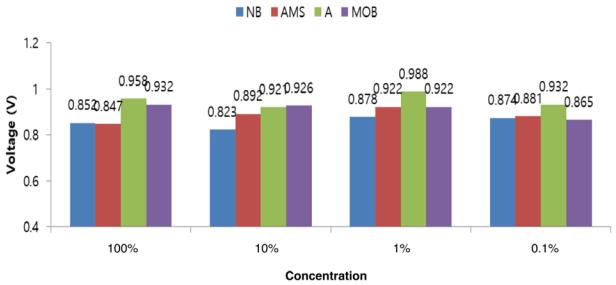


Figure 8: Symbiosis of SMO and MOB based on electrochemical behavior: voltage of PMFCs (petri dishes) with differing concentrations of NB & SMO A (red & blue) and AMS & MOB (green & purple).

As shown in Figure 8, the voltage measurements of the petri dishes with undiluted A and MOB (with concentrations of 100%) were greater than that of the petri dishes with undiluted NB and AMS, respectively. The same applies to the petri dishes with the bacteria/culture-media with concentrations of 10%. In short, the use of SMO A and MOB does increase the voltage of a PMFC when compared to the control groups with the NB and AMS culture media only.

However, there is not a clear relationship between the concentration of bacteria and voltage as the voltage measurements for all concentrations are not noticeably different. Yet, as mentioned above, higher concentrations of SMO A and MOB increase the voltage of the cell.

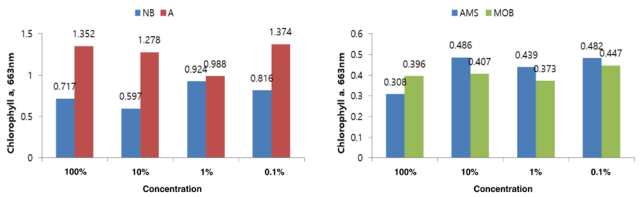


Figure 9: Symbiosis of SMO and MOB based on plant growth: chlorophyll a content in cabbage leaves grown in the agar medium treated with differing concentrations of NB & SMO A (left) and AMS & MOB (right).

As shown in Figure 9, chlorophyll a content in the cabbage leaves treated with SMO A was always higher than in the cabbage leaves treated with NB culture medium only. The difference in the chlorophyll content of the cabbage leaves treated with SMO A and NB culture medium only was the greatest for the undiluted, original solutions with concentrations of 100%. Chlorophyll a content in the cabbage

leaves treated with MOB was greater than in the cabbage leaves treated with AMS culture medium only for the undiluted, original solutions with concentrations of 100%. For all other concentrations, chlorophyll content was greater in the container with AMS medium only.

These results show that SMO A always had a positive effect on plant growth in terms of chlorophyll a content regardless of the concentration of the bacteria used. This is because the cabbage leaves treated with NB only were acting as the control group, and the cabbage leaves treated with SMO A along with NB displayed higher chlorophyll content. For similar reasons, the results show that MOB had a positive effect on plant growth in terms of chlorophyll a concentration but only when the solutions with the bacteria are not diluted.

The concentration of the bacteria did not seem to have a significant relationship with plant growth. Instead, the presence in itself, regardless of the concentration, of the bacteria (SMO A and MOB) affected plant growth. SMO A aided plant growth, while MOB hindered plant growth when it was diluted. In short, SMO A could be used at all concentrations and MOB should only be used without dilution.

Taking these results into consideration, the undiluted samples of SMO A and MOB were used in the subsequent experiment with paddy soil.

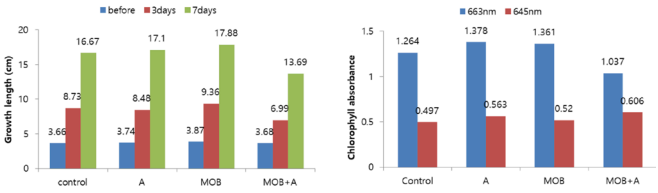
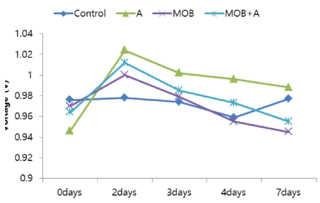


Figure 10: Height and chlorophyll content of the rice plants for varying combinations of SMO A and MOB (*Methylomonas sp.*).

The chlorophyll content and the height of the rice plants were monitored on days 0, 3, and 7 after treating them with SMO A and MOB, as shown in Figure 10. For both plant height and chlorophyll content, the two bacteria seemed to enhance the growth of rice plants when treated separately, with taller rice plants in their respective containers. However, rice plants showed slow growth when treated with both. This was likely due to the over-supply of nitrogen and other nutrients from the oxidation processes of the two bacteria, or the competition between SMO A and MOB.[8]

The rice plants did not grow best when both SMO A and MOB were present in the cell; it was when only MOB was added to the cell. But, all of the experimental conditions allowed for the growth of the rice plants, and none completely hindered the growth of the plants.

As shown in Figure 11, voltage peaked on day 2 for most PMFCs, with the SMO A container showing the largest increase of 0.946 V to 1.024 V. It can therefore be concluded that SMO A's oxidation process is the most rapid, having increased in conductivity the fastest and resulting in lower internal resistance compared to the control group on day 2. Containers with MOB seemed to decrease in voltage measurements rapidly, indicating that the soil did not have



	Control	A	MOB	MOB+A
0days	0.976	0.946	0.97	0.964
2days	0.978	1.024	1.000	1.012
3days	0.974	1.002	0.979	0.985
4days	0.959	0.996	0.955	0.973
7days	0.977	0.988	0.945	0.955

Figure 11: Change in voltage over time in rice plant containers (cells) with different combinations of SMO and MOB.

sufficient methane that can be oxidized. As determined in previous experiments, the container with both MOB and SMO A had higher voltage readings than that of the MOB-only PMFC. However, the container with SMO A was still at a higher voltage than the one with MOB and SMO A, disproving the initial hypothesis. This can be explained using the results from the previous experiment. As plant growth was deterred with the use of SMO A and MOB together, root exudates would have been insufficient for SMO A to oxidize, ultimately leading to smaller decreases in internal resistance (smaller net electrical output).

Conclusion

The aim of this study was to implement aerobic MOB in existing plant microbial fuel cell systems to increase their electrical efficiency. In regard to power output, *Methylomonas sp.* (MOB) and SMO A extracted from local rice fields increased the power output of the PMFC by 10.2% when used together. In regard to plant growth, it was observed that both SMO A and MOB enhanced the growth of rice plants, and that SMO A also aided the growth of cabbage. Treating rice plants with both bacteria prevented proper growth, likely a result of excess nutrients in the soil provided by the MOB and SMOs.

Thus, the synergism of MOB and the existing PMFC model is promising in that MOB enhances the power output when they are in an appropriate ratio with the soil microorganisms in the PMFC. Furthermore, MOB also improves the growth of plants in the PMFC, further enhancing the power output of the cell by helping the plant produce more organic material for the soil microorganisms to oxidize. However, the ratio of MOB to SMO may change as the two species proliferate in the paddy soil, possibly limiting electrical performance and plant growth over time. To prevent the addition of MOB to PMFC from counteracting its purpose of enhancing the cell's performance, more research must be conducted to find a way to control the ratio of SMO and MOB in a paddy field.

Acknowledgements

We would like to thank our research supervisor, Mr. Stuart Trivino, for his continued support. We would also like to thank our research mentor, SungKuk Kim at Nature Science Lab.

References

1. Yang, Y. S.; Lee, S. J.; Kim, A. IEA *World Energy Outlook 2018. World Energy Market Insight* 2018.
2. Chen, S.; Smith, A. L. Methane-Driven Microbial Fuel Cells Re cover Energy and Mitigate Dissolved Methane Emissions from Anaerobic Effluents. *Environmental Science: Water Research & Technology* 2018, 4 (1), 67–79.
3. Han, S.-K. Microbial Fuel Cells: Principles and Applications to En vironmental Health. *Korean Journal of Environmental Health Sci ences* 2012, 38 (2), 83–94.

DOI: 10.36838/v4i4.17

4. Sophia, A. C.; Sreeja, S. Green Energy Generation from Plant Microbial Fuel Cells (PMFC) Using Compost and a Novel Clay Separator. *Sustainable Energy Technologies and Assessments* 2017, 21, 59–66.
5. Koroglu, E. O.; Yoruklu, H. C.; Demir, A.; Ozkaya, B. Scale-up and Commercialization Issues of the MFCS. *Microbial Electrochemical Technology* 2019, 565–583.
6. Mirzababaei, J.; Chuang, S. La0.6Sr0.4Co0.2Fe0.8O3 Perovskite: A Stable Anode Catalyst for DIRECT Methane Solid Oxide Fuel Cells. Catalysts 2014, 4 (2), 146–161.
7. Krishnamoorthy, R.; Kwon, S.-W.; Kumutha, K.; Senthilkumar, M.; Ahmed, S.; Sa, T.; Anandham, R. Diversity of Culturable Methylotrophic Bacteria in Different Genotypes of Groundnut and Their Potential for Plant Growth Promotion. 3 Biotech 2018, 8 (6).
8. Larmola, T.; Leppanen, S. M.; Tuittila, E.-S.; Aarva, M.; Merila, P.; Fritze, H.; Tiirola, M. Methanotrophy Induces Nitrogen Fixation during Peatland Development. Proceedings of the National Academy of Sciences 2013, 111 (2), 734–739.

Author

SeungHan Ha is a senior at Chadwick International in Incheon, South Korea. He enjoys studying interdisciplinary areas in science, such as chemical engineering and environmental chemistry.

Daniel Seungmin Lee is a senior at Chadwick International in Incheon, South Korea. He enjoys exploring applications of physics in various fields, such as biochemistry.

IJHSR
ijhighschoolresearch.org
■ RESEARCH ARTICLE

Sheetal Kavach: Hybrid Cooling Jacket for Healthcare Workers in India

Vedant Singh
DPS RK Puram, New Delhi -110016, India; vedant.vasantvalley@gmail.com

ABSTRACT: A novel Hybrid Cooling Jacket (HCJ) using dry ice (solid CO_2) and ventilation fans has been developed for health care workers in India, required to work in hot and humid weather conditions. The goal of this HCJ was to decrease the related heat stress and heat shock associated with working in these conditions. The effectiveness of HCJ has been validated by trials with people in actual working conditions observed during the summertime.

Earlier developed personal cooling systems, used phase change materials (PCM) to absorb the body heat, however these had limitations like high weight, low thermal comfort, shorter duration of cooling (less than one hour), etc. The reason for low duration of cooling of the PCM based suits was the low latent heat of vaporization and hindrance of sweat evaporation.

The physical properties of a number of PCMs were studied and it was found that dry ice with its very high heat of sublimation can offer cooling which lasts 3-4 times longer and sweat produced by the body can be used to cool itself by evaporation. Through this jacket the cooling duration increased up to 4 hours with very little weight addition.

Trials performed on real people simulating similar ambient conditions with moderate activity, showed that skin temperature was maintained below 35.1˚C. Total sweat production was reduced by 21% and heart rate was reduced by an average 14 beats/min. The perceived thermal comfort parameters like thermal comfort votes, wetness sensation, and perceived exertion improved significantly.

KEYWORDS: Personal Cooling vest, Phase change material, Ventilation fans based suits, Air ventilation clothing, Dry ice jacket.

■ Introduction

Health care workers in India work for long hours and have to travel to remote places in hot and humid environments for vaccination and other medical duties, subjecting them to heat stress.[1] During the COVID-19 pandemic period, with the addition of personal protection equipment (PPE) suits, the problem of dissipating the heat has increased necessitating some cooling techniques for workers.

Personal protective clothing as a cooling solution has been attempted in past to increase body heat dissipation through evaporation, conduction, and convection. Phase change material (PCM) cooling vests are worn by workers outdoors.[2] The PCMs have high energy storage in the form of latent heat, have the advantage of absorbing the heat from the worker's body.[3] These kind of systems are known to be the cheapest, most portable, and the simplest to be worn.[4,5] The performance of the PCM cooling vests have been assessed in previous studies.[4-8] However, their effectiveness is constrained in hot humid environments due to reduced moisture transport and condensation.[9-12]

Another solution includes hybrid cooling suits/vests, which combine PCMs and battery-operated fans to enhance evaporative cooling and convective heat loss.[11-13] The PCMs absorb heat from the human body, while the fans enhance the sweat evaporation of through forced ventilation. However, most of these solutions have limitations like high weight, low thermal comfort, hindrance in the mobility, and most importantly, very short cooling duration due to the low latent heat of vaporization of the materials used for cooling (< 1 h).

Hybrid cooling suits/vests, which combine PCMs, and battery-operated fans are one of the most effective pathways to dissipate metabolic body heat to lower the core and skin temperatures with limitations as mentioned above.

To address these limitations, the physical properties of a number of phase change materials were studied and it was found that dry ice, with its very high heat of sublimation, can offer cooling which lasts 3-4 times longer. Moreover, dry ice is more practical to use as it melts into a gas instead of a liquid, leaving minimal cleanup. Additionally, the proposed solution was designed to use the sweat produced by the body to cool the body by evaporation.

Thus, a novel solution in the form of a hybrid cooling jacket using dry ice as phase change material for cooling and forced ventilation through fans for sweat evaporation was developed. Through this jacket the cooling duration increased by up to 4 hours with very little weight increase. The jacket was carefully tailored to allow an efficient air circulation inside the jacket by the ventilation fans with a discharge rate of 10 L/s to have a mean velocity in the range of 0.7-1.0 m/s for effective cooling. The position of fans and dry ice bags was such that relatively cool air reaches the most critical areas of body. To release the sublimated gas CO_2, two air vents were created at the crotch area of the jacket and PPE suit. The dry ice pouches were attached to jacket using 4-inch Velcro fasteners for easy replacement.

■ Methods

Hybrid cooling Jacket (HCJ):

Figure 1, Shows the newly designed hybrid cooling jacket

DOI: 10.36838/v4i4.18

made of polyester incorporates two dry ice packs (totaling 1.5 kg) placed evenly on the abdomen-chest and back areas and a pair of ventilation fans on the lower back area.

The ventilation fan is powered by a 12 V lithium battery which has a maximum air flow rate of 10 L/s and can operate for up to 6 hours. The weight of the ventilation unit (including a battery and a pair of ventilation fans) is 0.350 kg. The total weight of the cooling jacket is 1.95 kg. The HCJ mainly covered the wearer's torso, with a total covered body area of 0.7 m^2.

Figure 1: Hybrid Cooling Jacket worn over cotton t-shirt

Protocol of the experiment:

Trials were performed at an open terrace during hot summer weather in the month of July in Delhi. A total of 20 human trials (10 people X 2 tests each) at 36.0 ± 1 °C, RH = 55 ± 5% were carried out by testing subjects for 40-min walking on a treadmill at 5 km/h without a slope followed by a 20-minutes recovery by sitting under shade in the same place. Air temperatures and the relative humidity were measured using digital thermometers with a humidity sensor probe. The subjects wore HCJ over the regular cotton t-shirt and over that a PPE suit up to the subject's neck during both exercise and recovery periods and without HCJ and PPE suit in second set of trials. Heart rate was measured using heart rate band at every five minutes.

Measurement variables:

Physiological parameters: The parameters measured during the trials include heart rate and skin temperatures. Heart rate was measured using a heart band and temperatures using temperature probes attached to body with medical tape. Also, pre and post trials weights of the participants, their clothes and that of dry ice were measured using a weigh balance.

Perceptual Parameters: Participants were requested to report their thermal sensation votes (TSVs), thermal comfort votes (TCVs), skin wetness sensations (WSs) and ratings of perceived exertion (RPE) after every 10 minutes on a pre-defined standard questionnaire. TSVs were assessed using a continuous 9-point thermal sensation vote, ranging from –4 (very cold), –3 (cold), –2 (cool), –1 (slightly cool), 0 (neutral), +1 (slightly warm), +2 (warm), +3 (hot) to +4 (very hot).[14] TCVs were assessed by a 5-point thermal comfort scale, ranging from 0 (comfortable), +1 (slightly uncomfortable), +2 (uncomfortable), +3 (very uncomfortable) to +4 (extremely uncomfortable).[14] WSs were evaluated using a continuous 5-point scale ranging from 0 (neutral), +1(slightly wet), +2 (wet), +3 (very wet) to +4 (extremely wet). RPE was assessed using the 15-point Borg scale[15] (i.e., 6- 'no exertion', 7 & 8- 'extremely light', 9 & 10- 'very light', 11 & 12- 'light', 13 & 14- 'somewhat hard', 15 & 16- 'hard', 17 & 18- 'very hard', 19- 'extremely hard' and 20- 'maximal exertion'.

Calculations:

The mean skin temperature (T skin) and Sweat evaporation (SWevap) during the trial were calculated using equations as shown below: [9-13]

T mean skin = 0.25 (T chest + T scapula + T abdomen + T lumbar)

SWp = Wnude, start – Wnude, post–trial

SWcloth = Wclothed, start – Wclothed, post–trial

SWevap = SWp - SWcloth

where, Wnude, start, and W nude, post-trial is the nude body weight before and after the trials, and Wclothed, start and Wclothed, post-trial is the clothed body weight before and after the trials.

Results and Discussion

All participants successfully completed 20 trials. No significant differences were found between the HCJ and NHCJ in the baseline heart rate, mean skin temperatures, and perceptual responses prior to starting the trials ($p>0.05$)..

The total amount of sublimated dry ice in HCJ was 401 g over the trial duration and hence, the total cooling capacity of HCJ over 1 h trials was 61W (enthalpy of sublimation of dry ice: 571.3 kJ/kg; cooling capacity = (sublimated dry ice mass x enthalpy) / time duration).

Physiological parameters:

Figure 2 shows variations of heart rates, and mean skin temperatures in HCJ and NHCJ. Significant differences were found in the heartbeat rate between HCJ and NHCJ throughout the trial ($p< 0.001$). Heart rate was reduced by average 14 bpm and a maximum of 20 bpm in HCJ compared with NHCJ during the trials ($p<0.001$). With regard to torso/mean skin temperatures also significant lower temperatures were noticed in HCJ to NHCJ during the trial ($p< 0.001$).

The cooling jacket successfully prevented the wearers' mean skin temperatures from rising above 35.1 °C, whereas the mean skin temperature in NHCJ crossed 36.5 °C at the 30th min of the trial and it stayed above 36.5°C for the remaining period. The maximum mean skin temperatures in NHCJ was registered at the 30th min, i.e., 36.5 °C.

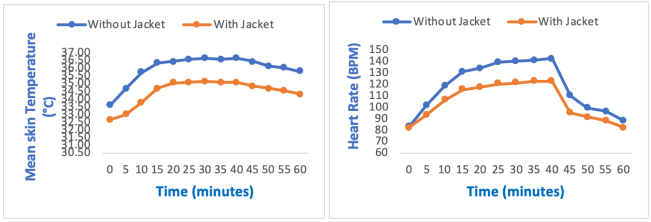


Figure 2: Variation of Mean Skin temperature and Heart Rate with time

Sweat production, sweat evaporation rate :

The total sweat production in HCJ and NHCJ during the duration of trials was 368 and 466 g/h, respectively. The total sweat evaporated over the trials in HCJ and NHCJ was 152 g and 193 g respectively. Cooling caused by sweat evaporation while wearing HCJ over one-hour trials was 104 W (enthalpy

of evaporation of water: 2460 kJ/kg; cooling capacity = (sweat evaporated x enthalpy) / time duration)

Perceptual responses:

Figure 3 shows the time course of perceptual responses including TSVs, TCVs, WSs, and RPE in the two studied test scenarios. Noticeable differences were found in all 4 perceptual indicators between HCJ and NHCJ throughout the trial. The maximum TSVs in the two test scenarios were registered at the 40th min, i.e., +3.4 (between 'hot' and 'very hot'), +2.3 (between 'warm' and 'hot') in HCJ and NHCJ respectively. Similarly, approaching the end of the 40-minute walking period, HCJ and NHCJ showed the highest TCV ratings, i.e., +2.5 (between 'uncomfortable' and 'very uncomfortable'), +1.7 ('uncomfortable') respectively. With regard to WSs, registered at the 40th min, i.e., +3.0 (very wet), +2.1 (wet) in HCJ and NHCJ. As for RPE values, registered at the 40th min, i.e., +14 (somewhat hard), +11.2 (light) in the test scenarios.

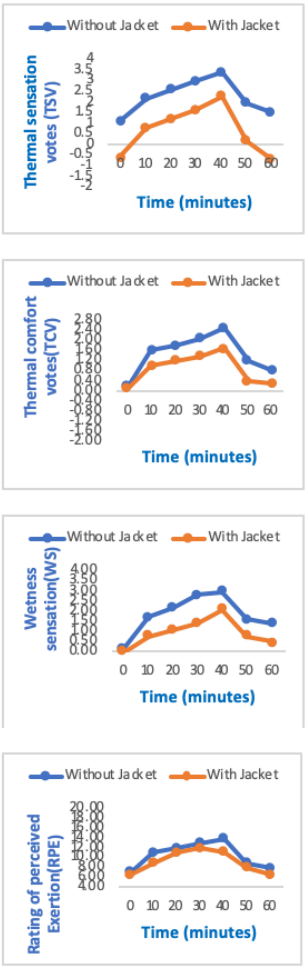


Figure 3: Variation of Preceptual parameters over the time of trial.

Discussion

This project developed a novel hybrid cooling jacket (HCJ) and examined its cooling effectiveness on for health care workers required to wear a PPE suit, while performing a moderate intensity activity in a hot and humid condition. A core temperature of 38.0 ˚C equivalent to skin temperature of 36-36.5˚C has been widely adopted as the threshold limit by reputable organizations such as WHO and ACGIH to protect outdoor workers. The jacket has successfully prevented the jacket wearers' mean skin temperatures from rising to temperatures greater than 35.1 ˚C wherein the wearers' mean skin temperature in NHCJ rose to 36.5 ˚C in about 30 min. In addition to the core temperature, heart rate and sweat rate are the other two most important physiological parameters to illustrate the exertion perceived by an individual person. The current findings indicated that HCJ, compared with NHCJ, depressed the heart rate by about an average 14 beats/min and maximum by 20 bpm. Moreover, the sweat rate in HCJ was considerably reduced by 21.1% which also reduced the dehydration rate correspondingly.

Participants had perceived lower TSVs, TCVs and WSs while wearing HCJ compared with those in NHCJ case ($P < 0.05$). This study had some limitations in that only young male participants took part in the trial. Individual factors like age, gender, body fat, and physical fitness might affect the body responses to personal cooling.

Conclusions

A portable hybrid cooling jacket was developed to mitigate heat stress of the health care workers performing in a hot and humid conditions (36 ± 1 ˚C, RH = 55 ± 5%, WBGT = 30.7 ˚C). Subsequently, trials were performed with real people simulating moderate activity conditions in hot and humid environments.

Acknowledgements

Foremost, I would like to acknowledge my parents for being highly supportive of this endeavor, for providing me with the required resources to develop the skills to take up this project, and for their general guidance during its evolution. I'd also like to thank Dr. Sandeep Chabra, Professor and Dr. Ashish Karnwal, HoD Mechanical Engineering, and KIET Ghaziabad for their support, guidance, and help in accessing the literature.

References

1. A.P. Chan, W. Yi, Heat stress and its impacts on occupational health and performance, Indoor Built environ. 25 (1) (2016) .
2. Johnson JK. Evaluation of four portable cooling vests for workers wearing gas extraction coveralls in hot environments. Graduate Theses and Dissertations. United States: University of South Flor ida; 2013
3. Sharma RK, Ganesan P, Tyagi VV, Metselaar HSC, Sandaran SC. Developments in organic solid–liquid phase change materials and their applications in thermal energy storage. Energy Convers Man age 2015; 95:193–228
4. Gao C, Kuklane K, Holmér I. Cooling vests with phase change materials: the effects of melting temperature on heat strain allevia tion in an extremely hot environment
5. Reinertsen RE, Færevik H, Holbø K, Nesbakken R, Reitan J, Røyset A, *et al*. Optimizing the performance of phase-change ma terials in personal protective clothing systems. Int J Occup Safety Ergon 2008; 14:43–53.
6. Bendkowska W, Kłonowska M, Kopias K, Bogdan A. Thermal manikin evaluation of PCM cooling vests. Fibres Textiles Eastern Europe 2010;18(1):78.
7. Gao C, Kuklane K, Holmér I. Cooling effect of a PCM vest on a thermal manikin and on humans exposed to heat. In: 12th International conference

DOI: 10.36838/v4i4.18

8. Zhao M, Gao C, Wang F, Kuklane K, Holmer I, Li J. The tor so cooling of vests incorporated with phase change materials: a sweat evaporation perspective. Textile Res J 2012; 83:418–25.
9. S.Á. Arngrïmsson, D.S. Petitt, M.G. Stueck, D.K. Jorgensen, K.J. Cureton, cooling vest worn during active warm-up improves 5-km run performance in the heat, J. Appl. Physiol. 96 (2004) 1867–1874
10. M. Zhao, C. Gao, F. Wang, K. Kuklane, I. Holmér, J. Li, the tor so cooling of vests incorporated with phase change materials: a sweat evaporation perspective, Text. Res. J. 83 (2013) 418–425.
11. W. Yi, Y. Zhao, A.P. Chan, Evaluating the effectiveness of cooling vest in a hot and humid environment, Ann. Work Expo. Health 61 (2017) 481–494.
12. D. Lai, F. Wei, Y. Lu, F. Wang, Evaluation of a hybrid personal cooling system using a manikin operated in constant temperature mode and thermoregulatory model control mode in warm condi tions, Text. Res. J. 87 (2017) 46–56
13. W. Song, F. Wang, the hybrid personal cooling system (PCS) could effectively reduce the heat strain while exercising in a hot and mod erate humid environment Ergonomics 59 (2016) 1009–1018
14. ISO, ISO10551, Ergonomics of the Thermal Environment—As sessment of the Influence of the Thermal Environment Using Subjective Judgement Scales, International Organization for Stan dardization, Geneva, 19
15. G.A. Borg, Psychophysical bases of perceived exertion, Med. Sci. Sports Exerc.14 (5) (1982) 377–381

Author

Vedant Singh is a junior at DPS RK Puram, New Delhi India. He enjoys studying math, physics, and language analysis. He is bronze medalist in International Olympiad of Linguistics and is interested in pursuing his major in Computer Science with a minor in Linguistics.

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Match Point: Predicting Outcomes of Hypothetical Tennis Matches Between Top 10 Ranked Players

Cayden J. Tu
Saint Ignatius College Preparatory High School, 2001 37th Ave. San Francisco, CA 94116, U.S.A.; cayden.y.tu@gmail.com

ABSTRACT: Tennis' popularity, along with its global viewership, is increasing. However, there has not been an abundance of analysis surrounding predictions of player matchups or a focus on significant statistics affecting game play in tennis. This paper describes a program that can determine the probability of a player winning against his opponent based upon the list of selected top-performing players. Moreover, this paper provides the logistic regression model used to calculate these winning probabilities; this summary demonstrates which variables were significant in predicting the outcome of a tennis match. For data, players were selected from the top 10 performers in men's tennis from 2015-2019. The individual and match statistics for the top ten players were utilized in developing this model; data was taken from the Grand Slam Tournaments.

The model created is a logistic regression model. A logistic regression model was deemed to be the most optimal method as it is designed to predict probabilities for a binary outcome. This model was successfully able to create predictions for the percentage that any of these players was able to beat any other player that was in the dataset. Overall, this model was successfully able to gather information about which players would have the highest chance of winning a tennis match and the most important factors that lead to this outcome.

KEYWORDS: Data Modeling; Probability and Statistics; Logistic Regression; Sports Analytics; Tennis Modeling.

Introduction

From different styles of play to ever changing stroke techniques, tennis is a sport that has evolved over time. Despite the various approaches throughout the past century, the rules of tennis have remained consistent. Although tennis has been a sport played for over 120 years, there has not been extensive research analyzing tennis statistics, especially when compared to research in other major American sports that are surrounded with an abundance of data science.[1] Even though tennis can be a team sport, competitive tennis is more often played as a single player rather than as a doubles partnership. The analysis proposed may more greatly benefit single players as many factors will not be affected by a teammate and are solely caused by their individual performance. The results may individually show the importance of each statistic to a tennis player.

In tennis, performance statistics are one of the most important aspects of a player's game. Strategy develops an individual's performance in various categories and contributes to the win or loss of a match. While statistical prediction models for tennis outcomes have been conducted for a few decades, each model has varying levels of accuracy.[2] The goal of this paper is to determine how strategy affects the outcome of a tennis match. A logistic regression model using Python was developed to analyze match data and determine the probability of a win for each individual player when matched against other players.

The current data science research surrounding tennis strategy highlights which skills are necessary to help an athlete win a match and are largely used by sports bettors. In 2019, researchers in China developed data-driven models of "point-by-point performance for male tennis players" in Grand Slam tournaments to determine how different contextual variables could be used to predict point outcomes of players based on their opposition.[3] Point-by-point datasets give more detailed information about a single match, as it shows the scoring progression and the player who served.[2] Gollub matched 12,000 point-by-point strings of the players who won and the points scored to generate a prediction and the match progression.[2] This logistic regression model had a 76.2% accuracy.[2]

Additionally, Barnett and Clark demonstrated that it is possible to develop models to show how important certain advantages are to a player.[4] Cornman and colleagues built on this to compare logistic regression, random forests, support vector machine, and neural networks using data from online tennis datasets.[5] Their research indicated that logistic regression and support vector machine have the highest predictability with almost 70% accuracy.[5] Gu and Saaty compared subjective judgments by tennis experts concerning who was most likely to win 94 US Open tennis matches.[6] Their data modeling predicted a 15% greater accuracy rate than that of tennis experts, indicating the benefits to statistical modeling in this sport.[5]

A variety of methods are utilized in sports analytics, especially when analyzing large amounts of data to assist with predicting outcomes. These research studies demonstrate the benefits and drawbacks of various analytical methods. The method selected for determining the potential outcomes in top 10 player matchups was a logistic regression. This type of analysis was utilized to determine the probability of a player winning his match. This paper displays the process for determining the probability of top ranked players winning against other top ranked players. It also shows which statistics

DOI: 10.36838/v4i4.19

are most important to determine who will win a particular tennis match.

Methods

Data:

The data that was utilized to perform these regressions was from match statistics of top ranked players from recent years. Data was obtained through "Ultimate Tennis Statistics," which is a repository for tennis data compiled by tennis statisticians.[7] The specific data utilized to develop this model was taken from the top 10 male tennis athletes during each of the five years between 2015-2019. Due to the limitations of the website design, data was individually collected from each player. Since this process was manualized, only the top ten players' data was selected. This five-year period was selected as a purposive sample to examine the most recent five-year span of tennis data immediately prior to the COVID-19 pandemic (as 2019 was the last full tennis season). Data was not selected post 2019 to avoid non-sampling errors (or potentially skewed data sets) resulting from crowd capacity, limited opponents, and other COVID-19 related game restrictions.

For each of these players, match statistics were gathered from each of the selected player's matches in a Grand Slam tournament for that certain year. The Grand Slam Tournaments data from these performers was solely used because players were placed into the brackets of greatest difficulty during these tournaments compared to other tournaments. The statistics that Ultimate Tennis Statistics provide for individual matches include first serve percentage, first serve won percentage, first serve return won percentage, second serve won percentage, second serve return won percentage, ace percentage, break points saved percentage, break points won percentage, and double fault percentage.[7] For a given match, these statistics were given for both players. The difference was calculated between the values for the top ranked player and their opponent to serve as the independent variables in the logistic regression models. Therefore, each of these statistics in the dataset represent the difference between the top ranked player and their opponent. In total, 786 tennis matches were gathered across all of the top 10 tennis players' matches.

In the dataset, there were statistics of a match that were direct inverses of each other. For instance, a player's first serve win percentage could be found by subtracting the opponent's first serve return win percentage from one. Therefore, a player's first serve win percentage and his opponent's first serve return won percentage would have to equal one. Similarly, the same is true for second serve won percentage and opponent's second serve return won percentage, as well as break points saved percentage and opponent's break points won percentage. Due to the nature of these factors, only first serve won percentage, second serve won percentage, and break points won percentage were used.

Data was available for these top ten male players for both individual matches and season-wide aggregates. In addition to grabbing the matches, the season-wide data was also gathered. This data included the same independent variables that were used above. The season-wide stats were used to demonstrate an average estimate of the player's performance relative to other players during that season. After the logistic regressions were created, the season-wide averages were used to predict the probability of each player winning a certain matchup.

Statistical Design:

After gathering the data for each individual match, this data was used to create a logistic regression to predict a winner of a specific tennis match. Analyzing this data and observing the significance of each aspect of the game (i.e., first serve percentage won as it relates to the game outcome) eventually led to a model that uses the previously studied data to see the outcome of any match between top 10 players in the range that was discussed earlier. When listing the outcome, it showed the probability of one player beating the other.

Logistic regression was appropriate for several reasons. First, a logistic regression model was able to describe the statistics in a clear way. Through the regression model, prediction charts may be developed to demonstrate the likelihood that one player will win against another specific player as long as both player's individual data sets have been included in the modeling. By using a logistic regression model, it was evident which variables were statistically significant. By knowing this, one can produce probabilities for which players would win in future matchups, as one player might be better than the other player in a category that is essential for a win. The logistic regression model fits around this concept of significant data and uses a player's success in this essential data against given opponents to determine future outcomes of these matches. For the model to be a good fit, it is important to experiment with different combinations of independent variables to produce the best result.

This model was extremely helpful for predicting outcomes as it gave a precise probability of who would win in each match based on an individual's previous matches. Not only does this allow outside individuals to predict winners of matches, but it also allows individual players to focus on which skills or strategies require improvement to win more matches. If tennis players could see how these statistics can change the outcome of their game, they would benefit in their game strategies.

Results and Discussion

A logistic regression model was developed to determine the correlative effect between a player's individual game statistics (like first serve won percentage) and the likelihood of winning the match against a similarly skilled opponent. To develop this model, the dependent and independent variables were carefully defined so that the model would not confound data between a match win and the statistics that led to past wins.

First, the dependent variable in this model is a win variable where one represents a win while a zero represents a loss. Since the model is meant to be predictive, win was a logical dependent variable to select because this type of model anticipates outcomes of events which may only have two outcomes; in the instance of tennis, the outcomes are either

wins or losses.

To create the program, the outcome of the match was recorded as a binary number, either being one for a win or zero for a loss. For instance, if Roger Federer defeated Andy Murray during a match in 2019, the dependent variable on Federer's data would have an output of the number one to indicate that he had won, whereas Murray would have an output of the number zero to indicate a loss. The statsmodel.api package in Python was used to create this logistic regression model. Using the "Logit" function associated with this package and incorporating the data as outlined, a regression was developed.

The independent variables in this program were the performance statistics that were discussed in the previous section (and defined in the Appendix): First Serve Won Percentage, Second Serve Won Percentage, First Serve Percentage, Break Points Won Percentage, Ace Percentage, and Double Fault Percentage. These are independent variables because their values are used in order to determine the probability that a given tennis player will win a hypothetical match against another player.

Table 1 demonstrates the results of the logistic regression for the matches of the top 10 ranked male players in Grand Slam tournaments from 2015-2019. The table is organized to emphasize the coefficients for independent variables. The significant independent variables were First Serve Won Percentage, Second Serve Won Percentage, First Serve Percentage, and Break Points Won Percentage (Table 1). The variable's significance is based on P values, where the lower the P value is, the more significant it is in predicting the outcome of the match. The coefficient shows the direction and magnitude of the effect of the variable.

Table 1: Logistic regression results to determine a win or loss.

Variable	Coefficients	Standard Error	*z*	*P* value
1st Serve Won %	33.71	4.31	7.83	<.001[a]
2nd Serve Won %	18.14	2.70	6.72	<.001[a]
1st Serve %	10.75	3.31	3.25	.001[a]
Break Points Won %	4.38	1.01	4.34	<.001[a]
Double Fault %	-2.38	8.42	-0.28	.78
Ace %	-3.20	3.10	-1.03	.30

[a] Indicates significant value.

Many of these variables had extremely low *P* values, acceptable at the .001 level, meaning that they were all significant towards the outcome of the game. The sign (either positive or negative) of the coefficient explains how it affects the dependent variable. A positive coefficient would mean that increasing the value of the independent variable will also greatly increase the probability of a win. The magnitude of the coefficient shows how big of an effect it has. The larger the coefficient, the greater the effect of a one unit increase or decrease of the independent variable on the dependent variable. Moreover, the coefficients of several of these statistics were relatively high. The lowest coefficient was "Break Points Won Percentage" at 4.38.

Ultimately, the ace percentage and double fault percentage were both deemed insignificant in predicting the outcome of a tennis match between top players. One reason why these factors might not prove essential to match outcome could be because aces and double faults do not make up a large portion of points in a tennis match. When professionals play against each other, double faults are rare, and aces are uncommon. Although the ace percentage was insignificant, it may have had a negative coefficient because it could have been correlated with double faulting and, therefore, aggressive serving. When a tennis player serves more aggressively, he or she is more likely to have aces but also more likely to have double faults. For tennis professionals, the serve and return are necessary for the game play, so players always focus on these parts of the point. Professional players make sure they do not "give" their opponents a free point by double faulting. Further, aces are rare because players are trained not to miss the ball completely on a serve or to let the ball bounce multiple times on his side of the net; both of these are relatively simplistic errors and somewhat upcoming to professional play.

To produce prediction probabilities for hypothetical matches of all the top 10 players against each other, another logistic regression model was created using only the four significant variables with LogisticRegression from the sklearn.linear_model package in Python. After creating this model, the "predict_proba" function was used to produce prediction probabilities. This package was used to predict probabilities due to a lack of function in the statsmodel.api package to perform the same functions.

The data that was used to predict the outcome of the matches were the season statistics. To help calculate the predictions, the program used the differences between the season-wide statistics across the whole range of top players. The difference of the season-wide statistics were the data that was used in predicting outcomes of all hypothetical matches among the top 10 players from 2015-2019. Furthermore, only the significant independent variables discussed earlier were put into the regression model. Including the insignificant statistics would skew the results of this model if the program used independent variables that were not significant towards the outcome of the match.

By focusing on the study findings, readers can see the importance of individual strategy on a player's game outcomes. For example, by discovering how significant serves are to a game's outcome, along with the importance of aces, players can determine how aggressive to be when serving and which type might yield the best outcome in the game. Also, the "break points won" percentage shows the level of importance in winning games when the tennis player has a high advantage where he needs to win that point to secure the win in that game, giving a perception of the importance of "clutchness." It is important for players to be able to win breakpoints as every time they have a lead, they are able to finish it so they can save energy and the momentum swing of an opponent's comeback.

After this model was created, there were many compelling findings. The model makes player rankings more evident. One such finding was that Milos Raonic had a greater than 50% win probability against all players during the 2018 season.

DOI: 10.36838/v4i4.19

However, the player who had less than a 50% probability to win against all players was Martin Klizan in 2018. Raonic in 2018 has one of the highest first serve win percentages out of all the players while Klizan in 2018 had one of the lowest. This shows how the model recognizes the importance of first serve won percentage and how much it takes part in determining the outcome of the match. In addition to finding the best and worst player with the average predicted probability against all the other players, it was also found that the top 10 players with the highest average predicted probability against other players and the bottom 10 (Table 2). The probability matrix in the Appendix contains a player's probability of beating each player in the dataset. The probabilities come from the player, whose name is located in rows, and his chance of beating each player, whose name is also in the columns. However, due to efficiency and readability, the matrix uses a number in place of each name. The number correlates to the row number the player was on in the dataset. For instance, in the table, when looking at Rafael Nadal in 2019, the number that takes its place in the matrix table is the number 1 as he was the best performer in 2019, leading him to be the highest in the dataset. After building the matrix, the mean probability was calculated by averaging the values of each row (Appendix).

From these results, there were also some interesting cases that despite the player performing better in the statistics that were significant towards the outcome, they still lost. This can occur many times as tennis is also a game of luck and there can be situations where a player is performing better than usual but is not able to pull off a win. For example, since each set is a race to 6, a player could win the first two sets 6-0, then lose the next three 7-5 but overall have better performance statistics.

Table 2: Top 10 best and worst players used in statistical analysis.

Rank	Player and Year	Average Probability
1	Milos Raonic 2018	87.0%
2	Roger Federer 2018	85.4%
3	Milos Raonic 2015	84.2%
4	Roger Federer 2015	83.1%
5	Roger Federer 2017	80.8%
6	Rafael Nadal 2019	78.9%
7	Rafael Nadal 2017	78.6%
8	Novak Djokovic 2015	77.6%
9	Roger Federer 2019	77.5%
10	Roger Federer 2016	76.9%
41	Rafael Nadal 2015	29.9%
42	Gael Monfils 2019	28.1%
43	Dominic Thiem 2019	27.1%
44	Kei Nishikori 2016	24.7%
45	Alexander Zverev 2017	24.1%
46	David Ferrer 2015	21.3%
47	Kei Nishikori 2019	20.7%
48	Alex De Minaur 2019	19.2%
49	Andrey Rublev 2019	11.8%
50	Martin Klizan 2018	6.8%

Conclusion

This model was successful in its attempt to display which statistics were significant out of the list of selected independent variables. The regression showed that the significant variables with a low P value were First Serve Won Percentage, Second Serve Won Percentage, First Serve Percentage, and Break Points Won Percentage. There were many possible reasons why these were the significant values as the other insignificant values are mostly uncommon occurrences in professional tennis games. After examining the predictions created from the regression model, there were several intriguing results. One such result was that a player could have a better performance in the statistics of the significant values than his opponent but would still lose on occasion. One reason for this could be a player dominating in the first few of his sets but losing in total as his opponent would win all the other sets by a small amount. Even though the player still won more points in total, the opponent still won all the other sets which secured him a win.

This paper was written to help tennis athletes and avid tennis-watchers to understand which statistics are significant towards determining the outcome of a match. Also, this paper provides another perspective on the use of Python to simulate tennis matches. As tennis popularity trends upwards, there still has not been much research on data science surrounding this sport. Nevertheless, tennis enthusiasts have been curious about the creation of a program that could determine which professional tennis player would beat another. Many game strategies are based upon which shots matter for a win. After this research, it is clear how important it is to win games starting with the first serve as it makes up for a large portion of the game. This emphasizes the importance of being ready for a return right after a strong first serve. If a player hits a strong first serve, returning another strong shot will be essential for starting the match. Tennis attracts many players, and it is a game of necessary strategy, making data science an excellent choice for studying this field and how players can perform their best.

Future Directions :

While the results from the initial analysis seem promising to predict tennis outcomes, repeating this modeling with other populations would increase the reliability. In the future, this model could be used on a larger sample size. Having more match data would demonstrate whether this model is applicable to a greater population. Further, including female tennis players in the sample to see if the model can still accurately predict outcomes in more than just male athletes would produce an interesting complement to this research.

Additionally, the data used in this sample was from 2015 to 2019–before the COVID-19 pandemic. It would be interesting to see how the model fits games played during 2020 and beyond, in which there would be different conditions such as little to no crowds, increased safety protocols, and possible psychological stressors for the players.

If the model described here is found to be consistent and generalizable with a larger population, creating a program in which a user-interface allows someone to pick two players

within a range and have it output their probabilities would produce a practical data model.

Acknowledgements

The work represented in this paper was not funded by any outside source. The author would like to thank and acknowledge Nick Riccardi, Adjunct Professor and Data Scientist at Syracuse University, for his mentorship during this project. Further, the author acknowledges his gratitude to his parents for their immense support during late nights spent writing.

References

1. Bruce, M. G. L.; Aberdare, B.; Lorge, B. S. Tennis, 2021. Ency clopedia Britannica. https://www.britannica.com/sports/tennis (accessed 2021-08-23).
2. Gollub, J. Producing Win Probabilities for Professional Tennis Matches from any Score. Bachelor's Thesis, Harvard College, Cambridge, MA, 2017. http://nrs.harvard.edu/urn-3:HUL.InstRepos:41024787 (accessed 2021-08-23).
3. Cui, Y.; Liu, H.; Liu, H.; Gómez, M. A. Data-Driven Analysis of Point-by-Point Performance for Male Tennis Player in Grand Slams. Motricidade. 2019, 15 (1), 49-61. DOI: 10.6063/motricidade.16370
4. Barnett, T.; Clarke, S. R. Combining Player Statistics to Predict Outcomes of Tennis Matches. *IMA Journal of Management Mathematics*. 2005, 16, 113-120. DOI: 10.1093/imaman/dpi001
5. Cornman, A.; Spellman, G.; Wright, D. Machine Learning for Professional Tennis Match Prediction and Betting. *Computer Science*. 2017.
6. Gu, W.; Saaty, T. L. Predicting the Outcome of a Tennis Tourna ment: Based on Both Data and Judgments. *Journal of Systems Science and Systems Engineering.* 2019, 28 (3), 317-343. DOI: 10.1007/s11518-018-5395-3
7. Ultimate Tennis Statistics. https://ultimatetennisstatistics.com/ (accessed 2021-08-23)

Authors

Cayden Tu is is the class of 2024 at Saint Ignatius High School in San Francisco. At school, he serves in leadership roles in various STEM organizations and the newspaper. His previous work includes coding, robotics, and creating artificial intelligence models focusing on disaster response efficiency and combating misogynistic verbal abuse. He enjoys playing tennis, sports, and data science and conceived this project to merge his passions.

Appendix

Ace Percentage: The percentage of a player's serve, typically the first serve striking with maximum force, that is so powerful to which the opponent is not able to return the serve at all by missing the ball completely or letting it bounce twice.

Double Fault Percentage: In tennis, a tennis player has two opportunities to start a point when it is his turn to serve. Typically, professional tennis players do not miss these two opportunities as that would be a free point awarded to the opponent. The double fault percentage shows out of all the points the athlete played when he was serving, how many of those times did he miss the second opportunity, which is supposed to be much more conservative than the first serve, where he hits the ball into the net or out of the box.

Break Points Won Percentage: Break points describe the points in a tennis game where a player is one point away from winning the game. In typical scoring, the player will be at the highest point number, 40, and if he wins this final point, then he secures the game. This is especially important for players to be able to finish their games when in the lead.

First Serve Won Percentage: The percentage of a player being able to win the point off of the first serve. This indicates that a player is able to win most of his points during the first serve.

Second Serve Won Percentage: The percentage of a player being able to win the point off of the second serve. This is usually much lower than the first serve win percentage as a

Player Code Reference:

PlayerYear	Player_Code
Rafael Nadal 2019	1
Novak Djokovic 2019	2
Roger Federer 2019	3
Daniil Medvedev 2019	4
Dominic Thiem 2019	5
Stefanos Tsitsipas 2019	6
Kei Nishikori 2019	7
Andrey Rublev 2019	8
Alex De Minaur 2019	9
Gael Monfils 2019	10
Rafael Nadal 2018	11
Roger Federer 2018	12
Novak Djokovic 2018	13
Juan Martin Del Potro 201	14
Alexander Zverev 2018	15
Dominic Thiem 2018	16
Kevin Anderson 2018	17
Martin Klizan 2018	18
Marin Cilic 2018	19
Milos Raonic 2018	20
Roger Federer 2017	21
Rafael Nadal 2017	22
Novak Djokovic 2017	23
Grigor Dimitrov 2017	24
Jo Wilfried Tsonga 2017	25
Alexander Zverev 2017	26
Andy Murray 2017	27
David Goffin 2017	28
Milos Raonic 2017	29
Juan Martin Del Potro 201	30
Andy Murray 2016	31
Novak Djokovic 2016	32
Milos Raonic 2016	33
Roger Federer 2016	34
Rafael Nadal 2016	35
Kei Nishikori 2016	36
Juan Martin Del Potro 201	37
Nick Kyrgios 2016	38
Gael Monfils 2016	39
Stan Wawrinka 2016	40
Novak Djokovic 2015	41
Roger Federer 2015	42
Andy Murray 2015	43
David Ferrer 2015	44
Kei Nishikori 2015	45
Stan Wawrinka 2015	46
Rafael Nadal 2015	47
Tomas Berdych 2015	48
Richard Gasquet 2015	49
Milos Raonic 2015	50

DOI: 10.36838/v4i4.19

Probability Matrix:

Player_Code	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	50.0%	53.9%	52.5%	93.3%	95.6%	89.4%	97.2%	98.8%	97.5%	95.3%	86.6%	36.5%	84.7%	83.7%	94.2%	84.3%	69.3%	99.4%	84.6%	33.0%	46.2%	50.6%	93.5%	88.8%	76.6%
2	46.1%	50.0%	48.7%	92.2%	94.9%	87.8%	96.7%	98.6%	97.1%	94.6%	84.7%	33.0%	82.5%	81.4%	93.3%	82.1%	65.9%	99.3%	82.4%	29.6%	42.4%	46.7%	92.4%	87.1%	73.7%
3	47.5%	51.3%	50.0%	92.6%	95.1%	88.3%	96.9%	98.6%	97.2%	94.9%	85.4%	34.2%	83.3%	82.2%	93.6%	82.9%	67.1%	99.3%	83.2%	30.8%	43.7%	48.1%	92.8%	87.7%	74.7%
4	6.7%	7.8%	7.4%	50.0%	60.5%	38.0%	70.4%	84.4%	72.8%	59.1%	32.2%	3.9%	28.9%	27.4%	53.6%	28.4%	14.2%	91.6%	28.8%	3.3%	5.8%	6.8%	50.6%	36.6%	19.5%
5	4.4%	5.1%	4.9%	39.5%	50.0%	28.6%	60.8%	77.8%	63.5%	48.5%	23.6%	2.5%	20.9%	19.7%	43.0%	20.5%	9.7%	87.6%	20.8%	2.1%	3.8%	4.5%	40.2%	27.4%	13.5%
6	10.6%	12.2%	11.7%	62.0%	71.4%	50.0%	79.6%	89.9%	81.4%	70.2%	43.7%	6.3%	39.9%	38.1%	65.3%	39.3%	21.5%	94.8%	39.8%	5.4%	9.2%	10.8%	62.6%	48.5%	28.4%
7	2.8%	3.3%	3.1%	29.6%	39.2%	20.4%	50.0%	69.2%	52.9%	37.8%	16.5%	1.6%	14.4%	13.5%	32.7%	14.1%	6.4%	81.8%	14.4%	1.4%	2.4%	2.9%	30.2%	19.4%	9.0%
8	1.2%	1.4%	1.4%	15.6%	22.2%	10.1%	30.8%	50.0%	33.3%	21.2%	7.9%	0.7%	6.8%	6.3%	17.7%	6.7%	2.8%	66.5%	6.8%	0.6%	1.1%	1.3%	16.0%	9.5%	4.1%
9	2.5%	2.9%	2.8%	27.2%	36.5%	18.6%	47.1%	66.7%	50.0%	35.1%	14.9%	1.4%	13.0%	12.2%	30.2%	12.7%	5.7%	80.0%	13.0%	1.2%	2.2%	2.6%	27.8%	17.7%	8.1%
10	4.7%	5.4%	5.1%	40.9%	51.5%	29.8%	62.2%	78.8%	64.9%	50.0%	24.7%	2.7%	21.9%	20.6%	44.5%	21.5%	10.2%	88.2%	21.8%	2.3%	4.0%	4.8%	41.6%	28.5%	14.3%
11	13.4%	15.3%	14.6%	67.8%	76.4%	56.3%	83.5%	92.1%	85.1%	75.3%	50.0%	8.0%	46.1%	44.3%	70.9%	45.5%	26.1%	95.9%	46.0%	6.9%	11.7%	13.7%	68.4%	54.9%	33.9%
12	63.5%	67.0%	65.8%	96.1%	97.5%	93.7%	98.4%	99.3%	98.6%	97.3%	92.0%	50.0%	90.7%	90.0%	96.7%	90.5%	79.8%	99.7%	90.7%	46.1%	59.9%	64.0%	96.2%	93.3%	85.2%
13	15.3%	17.5%	16.7%	71.1%	79.1%	60.1%	85.6%	93.2%	87.0%	78.1%	53.9%	9.3%	50.0%	48.2%	74.0%	49.4%	29.3%	96.5%	49.9%	8.0%	13.4%	15.7%	71.7%	58.7%	37.5%
14	16.3%	18.6%	17.8%	72.6%	80.3%	61.9%	86.5%	93.7%	87.8%	79.4%	55.7%	10.0%	51.8%	50.0%	75.4%	51.2%	30.8%	96.8%	51.7%	8.6%	14.3%	16.7%	73.1%	60.4%	39.3%
15	5.8%	6.7%	6.4%	46.4%	57.0%	34.7%	67.3%	82.3%	69.8%	55.5%	29.1%	3.3%	26.0%	24.6%	50.0%	25.5%	12.5%	90.4%	25.9%	2.9%	5.0%	5.9%	47.0%	33.3%	17.3%
16	15.7%	17.9%	17.1%	71.6%	79.5%	60.7%	85.9%	93.3%	87.3%	78.5%	54.5%	9.5%	50.6%	48.8%	74.5%	50.0%	29.8%	96.6%	50.5%	8.2%	13.7%	16.0%	72.2%	59.3%	38.1%
17	30.7%	34.1%	32.9%	85.8%	90.3%	78.5%	93.6%	97.2%	94.3%	89.8%	73.9%	20.2%	70.7%	69.2%	87.5%	70.2%	50.0%	98.6%	70.6%	17.8%	27.6%	31.2%	86.1%	77.5%	59.1%
18	0.6%	0.7%	0.7%	8.4%	12.4%	5.2%	18.2%	33.5%	20.0%	11.8%	4.1%	0.3%	3.5%	3.2%	9.6%	3.4%	1.4%	50.0%	3.5%	0.3%	0.5%	0.6%	8.6%	4.9%	2.1%
19	15.4%	17.6%	16.8%	71.2%	79.2%	60.2%	85.6%	93.2%	87.0%	78.2%	54.0%	9.3%	50.1%	48.3%	74.1%	49.5%	29.4%	96.5%	50.0%	8.1%	13.5%	15.7%	71.7%	58.8%	37.6%
20	67.0%	70.4%	69.2%	96.7%	97.9%	94.6%	98.6%	99.4%	98.8%	97.7%	93.1%	53.9%	92.0%	91.4%	97.1%	91.8%	82.2%	99.7%	91.9%	50.0%	63.6%	67.5%	96.8%	94.3%	87.1%
21	53.8%	57.6%	56.3%	94.2%	96.2%	90.8%	97.6%	98.9%	97.8%	96.0%	88.3%	40.1%	86.6%	85.7%	95.0%	86.3%	72.4%	99.5%	86.5%	36.4%	50.0%	54.3%	94.4%	90.2%	79.3%
22	49.4%	53.3%	51.9%	93.2%	95.5%	89.2%	97.1%	98.7%	97.4%	95.2%	86.3%	36.0%	84.3%	83.3%	94.1%	84.0%	68.8%	99.4%	84.3%	32.5%	45.7%	50.0%	93.3%	88.5%	76.2%
23	6.5%	7.6%	7.2%	49.4%	59.8%	37.4%	69.8%	84.0%	72.2%	58.4%	31.6%	3.8%	28.3%	26.9%	53.0%	27.8%	13.9%	91.4%	28.3%	3.2%	5.6%	6.7%	50.0%	36.0%	19.1%
24	11.2%	12.9%	12.3%	63.4%	72.6%	51.5%	80.6%	90.5%	82.3%	71.5%	45.1%	6.7%	41.3%	39.6%	66.7%	40.7%	22.5%	95.1%	41.2%	5.7%	9.8%	11.5%	64.0%	50.0%	29.7%
25	23.4%	26.3%	25.3%	80.5%	86.5%	71.6%	91.0%	95.9%	91.9%	85.7%	66.1%	14.8%	62.5%	60.7%	82.7%	61.9%	40.9%	97.9%	62.4%	12.9%	20.7%	23.8%	80.9%	70.3%	50.0%
26	3.6%	4.2%	4.0%	34.9%	45.1%	24.7%	56.0%	74.1%	58.8%	43.6%	20.2%	2.1%	17.7%	16.7%	38.3%	17.4%	8.0%	85.2%	17.7%	1.8%	3.1%	3.7%	35.5%	23.6%	11.3%
27	5.8%	6.7%	6.4%	46.4%	57.0%	34.7%	67.3%	82.3%	69.8%	55.5%	29.1%	3.3%	26.0%	24.6%	50.0%	25.5%	12.5%	90.4%	25.9%	2.9%	5.0%	5.9%	47.0%	33.3%	17.3%
28	12.8%	14.6%	14.0%	66.6%	75.4%	55.0%	82.8%	91.7%	84.4%	74.3%	48.7%	7.6%	44.8%	43.0%	69.8%	44.2%	25.1%	95.7%	44.7%	6.6%	11.1%	13.0%	67.2%	53.6%	32.8%
29	30.2%	33.6%	32.4%	85.5%	90.1%	78.1%	93.5%	97.1%	94.2%	89.6%	73.4%	19.8%	70.2%	68.7%	87.2%	69.7%	49.4%	98.6%	70.1%	17.4%	27.1%	30.7%	85.8%	77.1%	58.6%
30	7.2%	8.4%	7.9%	52.0%	62.4%	39.9%	72.0%	85.4%	74.3%	61.0%	34.0%	4.2%	30.5%	29.0%	55.6%	30.0%	15.2%	92.2%	30.5%	3.6%	6.2%	7.4%	52.6%	38.5%	20.8%
31	14.2%	16.2%	15.5%	69.2%	77.6%	57.9%	84.4%	92.6%	85.9%	76.5%	51.7%	8.6%	47.8%	45.9%	72.2%	47.2%	27.5%	96.2%	47.7%	7.4%	12.4%	14.5%	69.8%	56.5%	35.4%
32	12.2%	14.1%	13.4%	65.6%	74.5%	53.9%	82.1%	91.3%	83.7%	73.4%	47.5%	7.3%	43.6%	41.9%	68.8%	43.1%	24.2%	95.5%	43.6%	6.3%	10.7%	12.5%	66.2%	52.4%	31.7%
33	35.3%	38.9%	37.7%	88.2%	92.1%	81.9%	94.8%	97.7%	95.4%	91.6%	77.7%	23.8%	74.9%	73.4%	89.7%	74.4%	55.2%	98.9%	74.8%	21.1%	31.9%	35.8%	88.5%	81.0%	64.0%
34	46.3%	50.2%	48.8%	92.3%	94.9%	87.8%	96.7%	98.6%	97.1%	94.6%	84.8%	33.2%	82.6%	81.5%	93.3%	82.2%	66.1%	99.3%	82.5%	29.8%	42.6%	46.9%	92.5%	87.2%	73.8%
35	8.3%	9.6%	9.1%	55.5%	65.6%	43.3%	74.8%	87.1%	77.0%	64.3%	37.2%	4.8%	33.6%	32.0%	59.0%	33.1%	17.2%	93.2%	33.6%	4.1%	7.1%	8.4%	56.1%	41.9%	23.3%
36	3.7%	4.4%	4.1%	35.8%	46.0%	25.4%	56.9%	74.9%	59.7%	44.6%	20.8%	2.1%	18.3%	17.2%	39.2%	17.9%	8.3%	85.7%	18.3%	1.8%	3.2%	3.8%	36.4%	24.3%	11.7%
37	16.2%	18.4%	17.6%	72.4%	80.1%	61.6%	86.3%	93.6%	87.7%	79.2%	55.4%	9.9%	51.5%	49.7%	75.2%	50.9%	30.6%	96.7%	51.4%	8.5%	14.2%	16.5%	72.9%	60.1%	38.9%
38	22.1%	24.9%	23.9%	79.4%	85.6%	70.1%	90.3%	95.6%	91.3%	84.8%	64.5%	13.9%	60.8%	59.0%	81.7%	60.2%	39.2%	97.8%	60.7%	12.1%	19.6%	22.5%	79.8%	68.8%	48.2%
39	9.1%	10.5%	10.0%	58.1%	68.0%	46.0%	76.8%	88.3%	78.8%	66.7%	39.7%	5.4%	36.1%	34.4%	61.6%	35.5%	18.8%	93.9%	36.0%	4.6%	7.9%	9.3%	58.7%	44.5%	25.2%
40	13.5%	15.4%	14.7%	68.0%	76.5%	56.5%	83.6%	92.1%	85.2%	75.4%	50.2%	8.1%	46.3%	44.5%	71.0%	45.7%	26.3%	96.0%	46.2%	7.0%	11.8%	13.7%	68.5%	55.0%	34.1%
41	47.7%	51.6%	50.2%	92.7%	95.2%	88.4%	96.9%	98.6%	97.2%	94.9%	85.5%	34.4%	83.4%	82.3%	93.6%	83.0%	67.3%	99.3%	83.3%	31.0%	43.9%	48.3%	92.9%	87.8%	74.9%
42	58.5%	62.2%	60.9%	95.2%	96.9%	92.3%	98.0%	99.1%	98.2%	96.7%	90.2%	44.8%	88.7%	87.9%	95.9%	88.4%	76.1%	99.6%	88.7%	41.0%	54.8%	59.0%	95.4%	91.9%	82.3%
43	12.2%	14.0%	13.4%	65.5%	74.5%	53.8%	82.0%	91.3%	83.7%	73.3%	47.4%	7.3%	43.6%	41.8%	68.7%	43.0%	24.2%	95.5%	43.5%	6.3%	10.6%	12.5%	66.1%	52.3%	31.6%
44	2.9%	3.4%	3.3%	30.5%	40.2%	21.1%	51.1%	70.1%	53.9%	38.8%	17.1%	1.7%	15.0%	14.0%	33.7%	14.6%	6.6%	82.4%	14.9%	1.4%	2.5%	3.0%	31.1%	20.1%	9.4%
45	11.8%	13.6%	12.9%	64.7%	73.8%	52.9%	81.5%	91.0%	83.2%	72.6%	46.6%	7.0%	42.7%	40.9%	67.9%	42.1%	23.5%	95.4%	42.6%	6.0%	10.3%	12.1%	65.3%	51.4%	30.9%
46	15.1%	17.2%	16.4%	70.7%	78.8%	59.6%	85.3%	93.0%	86.7%	77.7%	53.4%	9.1%	49.5%	47.6%	73.6%	48.9%	28.8%	96.5%	49.4%	7.9%	13.2%	15.4%	71.2%	58.2%	37.0%
47	5.2%	6.1%	5.8%	43.8%	54.3%	32.3%	64.9%	80.7%	67.5%	52.9%	26.9%	3.0%	23.9%	22.6%	47.3%	23.5%	11.3%	89.4%	23.9%	2.6%	4.5%	5.3%	44.4%	31.0%	15.8%
48	19.0%	21.5%	20.6%	76.0%	83.0%	65.9%	88.5%	94.7%	89.6%	82.2%	60.0%	11.7%	56.2%	54.4%	78.6%	55.6%	34.8%	97.3%	56.1%	10.2%	16.7%	19.3%	76.5%	64.6%	43.5%
49	17.1%	19.5%	18.7%	73.7%	81.2%	63.2%	87.2%	94.0%	88.4%	80.3%	57.1%	10.5%	53.2%	51.4%	76.5%	52.6%	32.1%	97.0%	53.1%	9.1%	15.1%	17.5%	74.2%	61.8%	40.6%
50	60.8%	64.5%	63.2%	95.7%	97.2%	93.0%	98.2%	99.2%	98.4%	97.0%	91.1%	47.2%	89.7%	89.0%	96.3%	89.4%	77.9%	99.6%	89.6%	43.4%	57.2%	61.4%	95.8%	92.6%	83.7%

Player_Code	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
1	96.4%	94.2%	87.2%	69.8%	92.8%	85.8%	87.8%	64.7%	53.7%	91.7%	96.3%	83.8%	77.9%	90.9%	86.5%	52.3%	41.5%	87.8%	97.1%	88.2%	84.9%	94.8%	81.0%	82.9%	39.2%
2	95.8%	93.3%	85.4%	66.4%	91.6%	83.8%	85.9%	61.1%	49.8%	90.4%	95.6%	81.6%	75.1%	89.5%	84.6%	48.4%	37.8%	86.0%	96.6%	86.4%	82.8%	93.9%	78.5%	80.5%	35.5%
3	96.0%	93.6%	86.0%	67.6%	92.1%	84.5%	86.6%	62.3%	51.2%	90.9%	95.9%	82.4%	76.1%	90.0%	85.3%	49.8%	39.1%	86.6%	96.7%	87.1%	83.6%	94.2%	79.4%	81.3%	36.8%
4	65.1%	53.6%	33.4%	14.5%	48.0%	30.8%	34.4%	11.8%	7.7%	44.5%	64.2%	27.6%	20.6%	41.9%	32.0%	7.3%	4.8%	34.5%	69.5%	35.3%	29.3%	56.2%	24.0%	26.3%	4.3%
5	54.9%	43.0%	24.6%	9.9%	37.6%	22.4%	25.5%	7.9%	5.1%	34.4%	54.0%	19.9%	14.4%	32.0%	23.5%	4.8%	3.1%	25.5%	59.8%	26.2%	21.2%	45.7%	17.0%	18.8%	2.8%
6	75.3%	65.3%	45.0%	21.9%	60.1%	42.1%	46.1%	18.1%	12.2%	56.7%	74.6%	38.4%	29.9%	54.0%	43.5%	11.6%	7.7%	46.2%	78.9%	47.1%	40.4%	67.7%	34.1%	36.8%	7.0%
7	44.0%	32.7%	17.2%	6.5%	28.0%	15.6%	17.9%	5.2%	3.3%	25.2%	43.1%	13.7%	9.7%	23.2%	16.4%	3.1%	2.0%	18.0%	48.9%	18.5%	14.7%	35.1%	11.5%	12.8%	1.8%
8	25.9%	17.7%	8.3%	2.9%	14.6%	7.4%	8.7%	2.3%	1.4%	12.9%	25.1%	6.4%	4.4%	11.7%	7.9%	1.4%	0.9%	8.7%	29.9%	9.0%	7.0%	19.3%	5.3%	6.0%	0.8%
9	41.2%	30.2%	15.6%	5.8%	25.7%	14.1%	16.3%	4.6%	2.9%	23.0%	40.3%	12.3%	8.7%	21.2%	14.8%	2.8%	1.8%	16.3%	46.1%	16.8%	13.3%	32.5%	10.4%	11.6%	1.6%
10	56.4%	44.5%	25.7%	10.4%	39.0%	23.5%	26.6%	8.4%	5.4%	35.7%	55.4%	20.8%	15.2%	33.3%	24.6%	5.1%	3.3%	26.7%	61.2%	27.4%	22.3%	47.1%	17.8%	19.7%	3.0%
11	79.8%	70.9%	51.3%	26.6%	66.0%	48.3%	52.5%	22.3%	15.2%	62.8%	79.2%	44.6%	35.5%	60.3%	49.8%	14.5%	9.8%	52.6%	82.9%	53.4%	46.6%	73.1%	40.0%	42.9%	8.9%
12	97.9%	96.7%	92.4%	80.2%	95.8%	91.4%	92.7%	76.2%	66.8%	95.2%	97.9%	90.1%	86.1%	94.6%	91.9%	65.6%	55.2%	92.7%	98.3%	93.0%	90.9%	97.0%	88.3%	89.5%	52.8%
13	82.3%	74.0%	55.2%	29.8%	69.5%	52.2%	56.4%	25.1%	17.4%	66.4%	81.7%	48.5%	39.2%	63.9%	53.7%	16.6%	11.3%	56.4%	85.0%	57.3%	50.5%	76.1%	43.8%	46.8%	10.3%
14	83.3%	75.4%	57.0%	31.3%	71.0%	54.1%	58.1%	26.6%	18.5%	68.0%	82.8%	50.3%	41.0%	65.6%	55.5%	17.7%	12.1%	58.2%	86.0%	59.1%	52.4%	77.4%	45.6%	48.6%	11.0%
15	61.7%	50.0%	30.2%	12.8%	44.4%	27.8%	31.2%	10.3%	6.7%	41.0%	60.8%	24.8%	18.3%	38.4%	29.0%	6.4%	4.1%	31.3%	66.3%	32.1%	26.4%	52.7%	21.4%	23.5%	3.7%
16	82.6%	74.5%	55.8%	30.3%	70.0%	52.8%	56.9%	25.6%	17.8%	66.9%	82.1%	49.1%	39.8%	64.5%	54.3%	17.0%	11.6%	57.0%	85.4%	57.9%	51.1%	76.5%	44.4%	47.4%	10.6%
17	92.0%	87.5%	74.9%	50.6%	84.8%	72.5%	75.8%	44.8%	33.9%	82.8%	91.7%	69.4%	60.8%	81.2%	73.7%	32.7%	23.9%	75.8%	93.4%	76.5%	71.2%	88.7%	65.2%	67.9%	22.1%
18	14.8%	9.6%	4.3%	1.4%	7.8%	3.8%	4.5%	1.1%	0.7%	6.8%	14.3%	3.3%	2.2%	6.1%	4.0%	0.7%	0.4%	4.5%	17.6%	4.6%	3.5%	10.6%	2.7%	3.0%	0.4%
19	82.3%	74.1%	55.3%	29.9%	69.5%	52.3%	56.4%	25.2%	17.5%	66.4%	81.7%	48.6%	39.3%	64.0%	53.8%	16.7%	11.3%	56.5%	85.1%	57.4%	50.6%	76.1%	43.9%	46.9%	10.4%
20	98.2%	97.1%	93.4%	82.6%	96.4%	92.6%	93.7%	78.9%	70.2%	95.9%	98.2%	91.5%	87.9%	95.4%	93.0%	69.0%	59.0%	93.7%	98.6%	94.0%	92.1%	97.4%	89.8%	90.9%	56.6%
21	96.9%	95.0%	88.9%	72.9%	93.8%	87.6%	89.3%	68.1%	57.4%	92.9%	96.8%	85.8%	80.4%	92.1%	88.2%	56.1%	45.2%	89.4%	97.5%	89.7%	86.8%	95.5%	83.3%	84.9%	42.8%
22	96.3%	94.1%	87.0%	69.3%	92.6%	85.5%	87.5%	64.2%	53.1%	91.6%	96.2%	83.5%	77.5%	90.7%	86.3%	51.7%	41.0%	87.5%	97.0%	87.9%	84.6%	94.7%	80.7%	82.5%	38.6%
23	64.5%	53.0%	32.8%	14.2%	47.4%	30.2%	33.8%	11.5%	7.5%	43.9%	63.6%	27.1%	20.2%	41.3%	31.5%	7.1%	4.6%	33.9%	68.9%	34.7%	28.8%	55.6%	23.5%	25.8%	4.2%
24	76.4%	66.7%	46.4%	22.9%	61.5%	43.5%	47.6%	19.0%	12.8%	58.1%	75.7%	39.9%	31.2%	55.5%	45.0%	12.2%	8.1%	47.7%	79.9%	48.6%	41.8%	69.0%	35.4%	38.2%	7.4%
25	88.7%	82.7%	67.2%	41.4%	79.2%	64.6%	68.3%	36.0%	26.2%	76.7%	88.3%	61.1%	51.8%	74.8%	65.9%	25.1%	17.7%	68.4%	90.6%	69.1%	63.0%	84.2%	56.5%	59.4%	16.3%
26	50.0%	38.3%	21.1%	8.2%	33.1%	19.1%	21.9%	6.6%	4.2%	30.0%	49.1%	16.9%	12.1%	27.8%	20.1%	4.0%	2.5%	21.9%	55.0%	22.5%	18.1%	40.8%	14.3%	15.9%	2.3%
27	61.7%	50.0%	30.2%	12.8%	44.4%	27.8%	31.2%	10.3%	6.7%	41.0%	60.8%	24.8%	18.3%	38.4%	29.0%	6.4%	4.1%	31.3%	66.3%	32.1%	26.4%	52.7%	21.4%	23.5%	3.7%
28	78.9%	69.8%	50.0%	25.6%	64.8%	47.0%	51.2%	21.4%	14.6%	61.5%	78.3%	43.3%	34.3%	59.0%	48.5%	13.9%	9.3%	51.3%	82.1%	52.1%	45.4%	72.0%	38.8%	41.6%	8.5%
29	91.8%	87.2%	74.4%	50.0%	84.4%	72.1%	75.3%	44.3%	33.4%	82.4%	91.5%	68.9%	60.3%	80.8%	73.3%	32.2%	23.4%	75.4%	93.2%	76.0%	70.7%	88.4%	64.7%	67.4%	21.7%
30	66.9%	55.6%	35.2%	15.6%	50.0%	32.5%	36.2%	12.7%	8.3%	46.5%	66.0%	29.3%	22.0%	43.8%	33.8%	7.9%	5.2%	36.3%	71.2%	37.1%	31.0%	58.2%	25.5%	27.8%	4.7%
31	80.9%	72.2%	53.0%	27.9%	67.5%	50.0%	54.1%	23.5%	16.1%	64.3%	80.3%	46.3%	37.1%	61.8%	51.5%	15.4%	10.4%	54.2%	83.8%	55.1%	48.3%	74.4%	41.6%	44.5%	9.5%
32	78.1%	68.8%	48.8%	24.7%	63.8%	45.9%	50.0%	20.6%	14.0%	60.4%	77.5%	42.2%	33.3%	57.9%	47.3%	13.3%	8.9%	50.1%	81.4%	50.9%	44.2%	71.0%	37.6%	40.5%	8.1%
33	93.4%	89.7%	78.6%	55.7%	87.3%	76.5%	79.4%	50.0%	38.7%	85.6%	93.2%	73.7%	65.7%	84.2%	77.6%	37.5%	27.9%	79.5%	94.6%	80.0%	75.3%	90.6%	69.8%	72.3%	25.9%
34	95.8%	93.3%	85.4%	66.6%	91.7%	83.9%	86.0%	61.3%	50.0%	90.5%	95.7%	81.7%	75.2%	89.5%	84.7%	48.6%	38.0%	86.1%	96.6%	86.5%	82.9%	94.0%	78.6%	80.6%	35.7%
35	70.0%	59.0%	38.5%	17.6%	53.5%	35.7%	39.6%	14.4%	9.5%	50.0%	69.1%	32.3%	24.6%	47.4%	37.0%	9.0%	5.9%	39.7%	74.0%	40.5%	34.1%	61.6%	28.3%	30.8%	5.4%
36	50.9%	39.2%	21.7%	8.5%	34.0%	19.7%	22.5%	6.8%	4.3%	30.9%	50.0%	17.4%	12.5%	28.6%	20.7%	4.1%	2.6%	22.6%	55.9%	23.2%	18.6%	41.8%	14.8%	16.4%	2.4%
37	83.1%	75.2%	56.7%	31.1%	70.7%	53.7%	57.8%	26.3%	18.3%	67.7%	82.6%	50.0%	40.6%	65.3%	55.2%	17.5%	11.9%	57.9%	85.8%	58.8%	52.0%	77.2%	45.3%	48.3%	10.9%
38	87.9%	81.7%	65.7%	39.7%	78.0%	62.9%	66.7%	34.3%	24.8%	75.4%	87.5%	59.4%	50.0%	73.4%	64.3%	23.8%	16.7%	66.8%	89.9%	67.6%	61.3%	83.3%	54.7%	57.7%	15.3%
39	72.2%	61.6%	41.0%	19.2%	56.2%	38.2%	42.1%	15.8%	10.5%	52.6%	71.4%	34.7%	26.6%	50.0%	39.6%	10.0%	6.6%	42.2%	76.0%	43.1%	36.6%	64.1%	30.5%	33.1%	6.0%
40	79.9%	71.0%	51.5%	26.7%	66.2%	48.5%	52.7%	22.4%	15.3%	63.0%	79.3%	44.8%	35.7%	60.4%	50.0%	14.6%	9.9%	52.8%	83.0%	53.6%	46.8%	73.2%	40.2%	43.1%	9.0%
41	96.0%	93.6%	86.1%	67.8%	92.1%	84.6%	86.7%	62.5%	51.4%	91.0%	95.9%	82.5%	76.2%	90.0%	85.4%	50.0%	39.3%	86.7%	96.8%	87.1%	83.7%	94.3%	79.5%	81.5%	37.0%
42	97.5%	95.9%	90.7%	76.6%	94.8%	89.6%	91.1%	72.1%	62.0%	94.1%	97.4%	88.1%	83.3%	93.4%	90.1%	60.7%	50.0%	91.1%	97.9%	91.4%	88.9%	96.3%	85.9%	87.3%	47.6%
43	78.1%	68.7%	48.7%	24.6%	63.7%	45.8%	49.9%	20.5%	13.9%	60.3%	77.4%	42.1%	33.2%	57.8%	47.2%	13.3%	8.9%	50.0%	81.3%	50.9%	44.1%	71.0%	37.6%	40.4%	8.1%
44	45.0%	33.7%	17.9%	6.8%	28.8%	16.2%	18.6%	5.4%	3.4%	26.0%	44.1%	14.2%	10.1%	24.0%	17.0%	3.2%	2.1%	18.7%	50.0%	19.2%	15.2%	36.1%	12.0%	13.3%	1.9%
45	77.5%	67.9%	47.9%	24.0%	62.9%	44.9%	49.1%	20.0%	13.5%	59.5%	76.8%	41.2%	32.4%	56.9%	46.4%	12.9%	8.6%	49.1%	80.8%	50.0%	43.2%	70.3%	36.7%	39.6%	7.8%
46	81.9%	73.6%	54.6%	29.3%	69.0%	51.7%	55.8%	24.7%	17.1%	65.9%	81.4%	48.0%	38.7%	63.4%	53.2%	16.3%	11.1%	55.9%	84.8%	56.8%	50.0%	75.7%	43.3%	46.2%	10.1%
47	59.2%	47.3%	28.0%	11.6%	41.8%	25.6%	29.0%	9.4%	6.0%	38.4%	58.2%	22.8%	16.7%	35.9%	26.8%	5.7%	3.7%	29.0%	63.9%	29.7%	24.3%	50.0%	19.6%	21.6%	3.4%
48	85.7%	78.6%	61.2%	35.3%	74.5%	58.4%	62.4%	30.2%	21.4%	71.7%	85.2%	54.7%	45.3%	69.5%	59.8%	20.5%	14.1%	62.4%	88.0%	63.3%	56.7%	80.4%	50.0%	53.0%	13.0%
49	84.1%	76.5%	58.4%	32.6%	72.2%	55.5%	59.5%	27.7%	19.4%	69.2%	83.6%	51.7%	42.3%	66.9%	56.9%	18.5%	12.7%	59.6%	86.7%	60.4%	53.8%	78.4%	47.0%	50.0%	11.6%
50	97.7%	96.3%	91.5%	78.3%	95.3%	90.5%	91.9%	74.1%	64.3%	94.6%	97.6%	89.1%	84.7%	94.0%	91.0%	63.0%	52.4%	91.9%	98.1%	92.2%	89.9%	96.6%	87.0%	88.4%	50.0%

DOI: 10.36838/v4i4.19

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

How A Rotating Orbital Tether Advances Space Travel

Ufuk Çetiner
American Robert College of İstanbul Arnavutköy, Kuruçeşme Cd. No:87, 34345 Beşiktaş/İstanbul; cetinerufuk1@gmail.com

ABSTRACT: The idea of an orbital tether was first proposed by the Russian scientist Konstantin Tsiolokovski. This study optimizes and develops on the idea of the Russian scientist. By combining different methods, this study shows how a momentum exchanging rotating orbital tether can be built for space travel and it compares this method of space travel to the conventional methods of space travel.

KEYWORDS: Momentum exchange; rotating; Skyhook; space travel; tether.

Introduction

As technology advances, traveling to other planets and colonizing them seem inevitable. Although there are companies that try to commercialize a space experience, what they offer is no more than five minutes of zero G. The main reason why space travel hasn't been commercialized yet is because of the lack of access to the Earth's orbit. It takes giant rockets filled with expensive chemical propellants to get into the Earth's orbit and then launch to outer space. This makes a rocket launch very expensive and dangerous which makes space travel unaffordable.

The Earth Orbiting Tether or the structure called Skyhook aims to reduce the cost of putting a satellite or a payload into low earth orbit. It allows payloads or satellites to be raised to a higher orbit in which then they can use a meager amount of propellant to accelerate to escape velocity. The structure is built with a strong fiber that makes up the tether which orbits around the Earth.

This study aims to optimize, improve, and build on the idea while also comparing it to the conventional ways of space travel. The technologies that are discussed in this study are based on existing technologies.

Importance of The Subject:

If this project were to be built, its impact on space exploration would be significant. It would significantly reduce the propellant needs in space missions, which would make these missions much cheaper compared to their predecessors. Since the tether can accelerate objects into space, it can do the opposite: decelerating incoming objects. This allows landing on places where it is hard to decelerate, for example, asteroids. Since there is no atmosphere to slow the objects down; the orbiting tether can do that job. This would allow asteroid mining which would advance humanity in resource gathering.

Discussion

Background Research:

The concept of a skyhook was first proposed by a Russian scientist named Konstantin Tsiolokovski. He wrote about the possibility of an object's withdrawing from the earth with the aid of a rigid structure (tower) that would extend upward at right angles to the earth's surface.[1] Artsutanov then developed his idea with a 36,000 kilometer stationary satellite.[1] Extending straight up and down from the satellite are the lengths of the cable.[1] The lower end of the cable is secured to the earth's surface and the upper end does not need to be anchored as if it were on a "skyhook."[1]

There hasn't been a skyhook sent into the atmosphere, but there have been a number of flight experiments exploring various aspects of the space tether concept in general. One of these being a draft proposal for a skyhook system for NASA under the name of HASTOL (Hypersonic Airplane Space Tether Orbital Launch).[2]

Proposed Orbital Tether Structure Operational Methodology:

In a momentum exchanging orbital tether, a long thin cable is placed in orbit and set to rotate around a center of mass. In one end of the tether a docking structure is present for an incoming payload to connect to the tether at a lower orbit. If the rotation of the tether is timed correctly with the payload's position, a specialized quadtrap can dock the payload to the tether (See Figure 1). The details of this structure will be discussed throughout this section. After the payload has successfully docked to the tether, the rotation of the structure will accelerate it to a higher orbit. The structure then releases the payload at a high orbit setting it on course. After the release of the payload the tether is off its operational orbit. In order to readjust its orbit to the operational one, the tether can catch an incoming payload to regain momentum. If there are no arriving payloads at that time the tether uses the ion boosters to readjust its course. The operational summary can be seen in Figure 2.

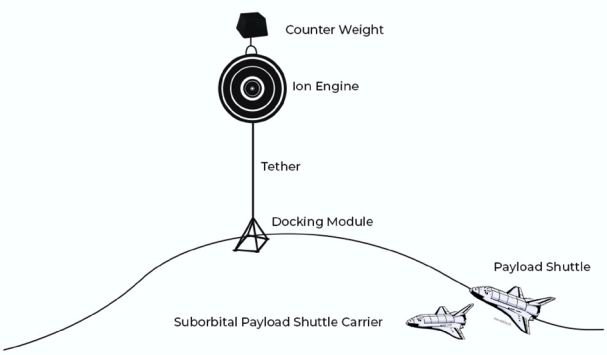


Figure 1: Momentum exchanging rotating orbital tether structure operation overview.

DOI: 10.36838/v4i4.20

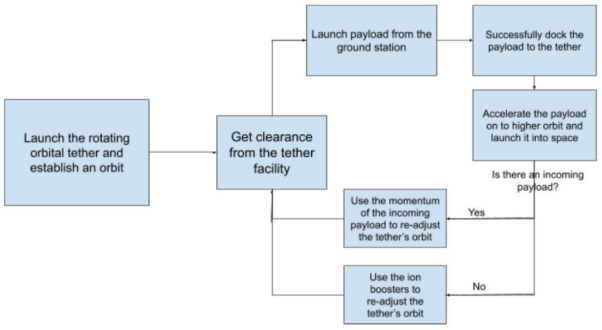


Figure 2: Operational summary of momentum exchanging orbital tether structure.

Ion Boosters:

After launching each payload into outer space, the structure loses momentum. The structure may slow down other incoming payloads to regain its momentum. However, if there are no incoming payloads at that time or if an emergency takes place the structure will need an auxiliary system to correct its orbital course. Ion boosters placed in the center of mass of the structure will help with the re-adjustment of the orbit at will. Ion booster's efficient use of fuel and electricity make modern spacecrafts travel farther, faster, and cheaper.[3] This structure will have seldom refueling since the engines on board need a small amount of fuel to continue the operation for long periods of time. In summary, the high fuel efficiency of ion boosters makes them a great choice for the proposed structure. For instance, the ion boosters in the Deep Space 1 mission used less than 159 pounds of fuel in over 16,000 hours of thrusting.[3]

Tether:

A tether, also known as a space tether in the context of space traveling, is a long cable which can be used for propulsion, momentum exchange, stabilization and altitude control, or maintaining the relative positions of the components of a large, dispersed satellite spacecraft sensor system.[4] The proposed structure will employ a rotating tether for momentum exchange, which essentially works by creating a controlled force on the end-masses of the system.

The tether in this structure will be exposed to intense amounts of tension so the fiber should be strong. A synthetic fiber called zylon is designed to withstand tensions like this. The fiber is like Kevlar that is used in bullet proof vests and will be used in this structure as the tether's main component.

Docking Station:

Docking to the momentum exchanging rotating orbital tether structure is a challenging task. Since the docking tip of the structure is traveling at high speeds, timing the docking sequence is paramount to the mission's success. However conventional docking procedures that are used in current space missions take up minutes.[5] To dock to this structure the aircraft will have no longer than seconds, so a new method of docking must be used. A method proposed by Soresen Kirk and his colleagues in a recent study will be taken in consideration for this research paper. At the tip of the tether a mechanism named "quadtrap" will be attached. Quadtrap is a rectangular mechanism that has two configurations: and open (See Figure 3). The mechanism quadtrap acts as a trap to guide the payload to a center point. In an open or deployed position, the capture mechanism has all four spreader bars translated at a maximum displacement from each other (See figure 3).[6] In a fully closed position the four corner nodes come together to close around a point.[6] In the middle position the payload has been securely captured and the docking window is met.[6]

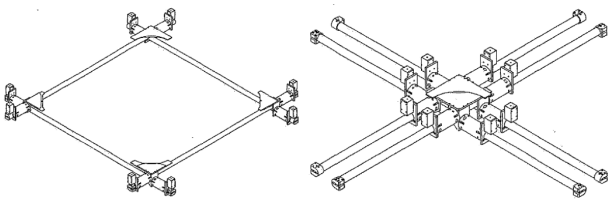


Figure 3: Quadtrap design open and closed configurations.[6]

Suborbital payload shuttle:

The docking tip of the momentum exchanging orbital tether dips down to earth around 80 to 100 kilometers above the sea level. This means that the payload cannot be delivered via jet engines since they can't reach such extreme heights. The proposed solution for this is a combination of both jet engines and rocket propulsion. The payload will be delivered via a suborbital space shuttle similar to the spacecraft "SpaceShipOne."[7] A payload shuttle will be specifically designed to help with the docking of the payload. The designed payload shuttle will be attached to a normal jet engine powered aircraft that will carry it to the maximum cruising altitude. From there the rocket-powered payload shuttle will launch itself in the air reaching the desired height and then glide back to safety with its wings; similar to the mission "Tier One" which involved the spacecraft "SpacesShipOne" and the carrier "White Knight"(See Figure 4). This design is very effective in reducing costs since the spacecraft is reusable and if the payload fails to dock the payload won't go to waste.

Figure 4: SpaceShipOne suborbital spacecraft attached to the carrier White Knight.[7]

Comparison to conventional methods of space travel:

All conventional space missions have large amounts of rocket fuel to get the payload to escape velocity. If you want to add more payload you need more rocket fuel. More rocket fuel means a heavier rocket, and heavier the rocket means

more rocket fuel. So, in all space missions, mass is arguably the biggest constraint. For each kilogram of payload mass, a significant amount of propellant is needed for altitude control, orbit transfers, momentum dumping, station-keeping, and a variety of other applications during the lifetime of the mission. Currently, the average cost to launch a kilogram of payload mass into Low-Earth Orbit (LEO) is $10,000.[8]

With the use of the momentum exchanging orbital tether that is being mentioned in this research heavier payloads can be launched into space with ease. Since large amounts of rocket fuel is not required to accelerate the payload to the escape velocity, larger amounts of the total mass can be dedicated to actual payload and not to rocket fuel. To make a comparison the previous Mars mission "Perseverance" will be examined.

Table 1: From Budget of Mars mission "Perseverance".[9]

Service	Cost
Spacecraft Development	2.2 billion dollars
Launch Services	243 million dollars
Operations and Scientific Analysis	300 million dollars
Ingenuity Mars Helicopter	80 million dollars
Total Cost of the Mission	2.7 billion dollars

As listed in Table 1, the launch services of the latest Mars mission took up about 9% of the whole budget. The latest Mars mission was expensive because it included state of the art equipment that's why the launch services seem like a small fee. However if a mission required just to send building materials to Mars the cost of sending something would be much more than the payload sent.

On the other hand for the momentum exchanging orbital tether the launch services solely depend on the reusable spacecraft (after the structure has established an orbit). Since the suborbital payload shuttle discussed in this paper was similar to the spacecraft used in the "Tier One" mission we can compare the costs of development, construction and operation. Although the costs of development, construction and operation for mission "Tier One" were not publicly shared it was estimated to be around 20 to 30 million dollars.[10] This means that this method of space travel can operate continuously with just 1% of money spent on one-time conventional launch service.

Conclusion

Space travel, while being a topic of high interest, is usually very difficult and expensive. This study has demonstrated that the solution to make space travel cheap and easy has already been found. With the correct combinations of currently existing technology a fully functioning momentum exchanging orbital tether can be built. Not only has this study shown that this design can work it also makes current methods of space travel irrelevant. This is because this method is highly reusable and a lot cheaper. Future research on this topic should include precise and accurate calculations of the instruments mentioned in this research.

Acknowledgements

I would like to thank my friends for their valuable opinions throughout this research.

References

1. Lvov, V. Sky-Hook: Old Idea. https://www.google.com/url?q=https://www.science.org/doi/pdf/10.1126/science.158.3803.946&sa=D&source=docs&ust=1643547005798064&usg=AOvVaw26BuNBwg_bMULespUmq9ba (accessed Jan 30, 2022).
2. Boeing. Hypersonic airplane space tether orbital launch (HASTOL). http://www.niac.usra.edu/files/studies/final_report/391Grant.pdf (accessed Jan 30, 2022).
3. Dunbar, B. Ion Propulsion: Farther, faster, cheaper. https://www.nasa.gov/centers/glenn/technology/Ion_Propulsion1.html (accessed Jan 30, 2022).
4. Cosmo, M. L. Tethers In Space Handbook. https://web.archive.org/web/20111027063548/http://www.nasa.gov/centers/marshall/pdf/337451main_Tethers_In_Space_Handbook_Section_1_2.pdf (accessed Jan 30, 2022).
5. Wertz, J. R.; Bell, R. Autonomous Rendezvous and Docking Technologies Status and Prospects. https://web.archive.org/web/20120425122952/http://microcosminc.com/analysis/spie03.pdf (accessed Jan 30, 2022).
6. Kirk, S. F. Design rules and analysis of a capture mechanism for rendezvous between a space tether and payload - NASA technical reports server (NTRS). https://ntrs.nasa.gov/citations/20070001536 (accessed Jan 30, 2022).
7. Wikipedia. SpaceShipOne. https://en.wikipedia.org/wiki/SpaceShipOne (accessed Jan 30, 2022).
8. Jones, H. W. The Recent Large Reduction in Space Launch Cost. https://ttu-ir.tdl.org/bitstream/handle/2346/74082/ICES_2018_81.pdf (accessed Jan 30, 2022).
9. Planetary Society. Cost of perseverance. https://www.planetary.org/space-policy/cost-of-perseverance (accessed Jan 30, 2022).
10. Wikipedia. Scaled composites tier one. https://en.wikipedia.org/wiki/Scaled_Composites_Tier_One#Funding (accessed Jan 30, 2022).

Author

Ufuk Çetiner is a sophomore student as of 2022 studying at the American Robert College of İstanbul. He runs the underwater robotics team in his school and plans to pursue a career in engineering.

DOI: 10.36838/v4i4.20

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Exosome-encapsulated miRNAs as Protective Agents against Huntington's Disease

Yejin Kong
Seoul Foreign School, Jongro-gu, Bukchon-ro, 03051, Seoul, Republic of Korea; yejin.kong.1118@gmail.com

ABSTRACT: Huntington's disease (HD) kills approximately 2.27 people per million every year and slowly makes everyday activities more difficult for those with the genetic mutation. Key symptoms of HD include chorea, forgetfulness, impaired judgment, personality changes, mood swings, and depression. HD is clinically diagnosed by the HD mutation in the huntingtin gene: overly repeated C-A-G nucleotide sequences. Abnormal sequencing in the gene causes abnormal protein synthesis in the huntingtin protein, the impact of which compounds over time as the proteins create more clumps in the brain, which is why HD is mostly diagnosed in middle-aged individuals - when the proteins have had enough time to cause symptoms. HD is caused by prolonged damage to the striatum caused by the abnormal HD gene and protein which makes brain cells eventually die. The link between miRNAs and HD has been studied before but this study takes a step further by examining how this link can be exploited to create a cure. Using three types of strategically selected miRNAs, this study showed that encapsulating the miRNAs in exosomes for delivery into heavily damaged neurons increased cell proliferation. The 6-hydroxydopamine (6-OHDA)-induced HD model was used to test the impact of three different miRNA types and showed that all three miRNAs increase cell proliferation. Overall, this research focuses on the potential of miRNA with exosome use in therapeutic treatment for blocking neuronal cell death of HD patients. As current techniques evolve, this research can be used for the development of personalized therapy for HD patients.
KEYWORDS: Molecular Biology, miRNA, Huntington Disease, 6-OHDA, Cell Death.

Introduction

HD is caused by a genetic mutation in the Huntingtin gene (IT15) where the C-A-G nucleotide sequence is overly repeated. This discovery has been named the "CAG repeat expansion" where the general population has around 20 repeats in this gene, those with clinically diagnosed HD have over 40. Everyone with this mutation will show symptoms at some point in their lives, but the reason why most patients are middle-aged when first diagnosed is because of the impact the genetic mutation has on huntingtin protein synthesis. The specifics of this protein's function have not been elucidated clearly yet, but scientists assume that the CAG repeat expansion causes abnormal protein function. These proteins are consistently produced over the span of a lifetime and form clumps in the brain causing both physical and neurological symptoms to worsen over time.[1] Although this phenomenon affects the entire brain, the striatum is most heavily impacted. HD does not have a cure at the moment partially because there simply is not enough information on the specifics of what happens to those with the mutation from a scientific standpoint.[2]

Retinoic acid is a small lipophilic molecule derived from vitamin A[3] that plays a key role in cell growth and differentiation.[4] In this experiment, it was used to differentiate the A172 Glioblastoma cell to mimic a neuron cell. Normally, retinoic acid is used to differentiate embryonic stem cells into motor neurons by altering both encoding RNA and miRNA expression[5] and this is the main objective of retinoic acid use in this experiment as well. Cell differentiation is the process in which a cell transforms into a more specialized type. This process changes a cell's shape, size, and energy requirements.[6] Since HD is a neurodegenerative disease that is characterized by the gradual and progressive loss of neurons,[7] an experiment aiming to find protective agents against the disease must take place using neurons, the cell type that's affected by the disease.

Dopamine is a neurotransmitter that sends signals between neuron cells and is an essential part in controlling voluntary movements. An alteration of the dopamine balance in the striatum specifically leads to pathological conditions like HD. Changes in the amount of dopamine and its receptors cause abnormal movements and cognitive deficits. Evidence also shows that an increased level of dopamine release induced chorea, a defining symptom of HD, while a reduction leads to akinesia, the inability to make any voluntary action.[8] Although the reason behind cell loss in HD is unclear, it may be explained by excessive glutamate release from cortical and thalamic terminals or an increased sensitivity of glutamate receptors.[9] Glutamate is another type of neurotransmitter and its receptor's function is controlled by the activation of dopamine receptors. This relationship suggests that an alteration in dopamine function and neurotransmission would significantly impact the motor and cognitive symptoms of HD.[9]

Using this relationship, this experiment uses 6-hydroxydopamine (6-OHDA), a form of dopamine that has been used in Parkinson's disease studies to induce neuronal damage. 6-OHDA causes massive destruction of neurons and experts have been focusing on 6-OHDA's intracellular mechanisms at the striatal level.[10] Therefore, 6-OHDA can be used to mimic

DOI: 10.36838/v4i4.21

increased mutated huntingtin (mHTT) expression while suppressing HTT phosphorylation at Ser421, a modification that would have protected it against mHTT accumulation.[11]

Among many of its functions, such as maintaining the division amongst micro and macromolecules, exosomes are also produced by almost every cell in the body and are shown to impact distant cell signaling.[12] Since they are a commonly produced mobile extracellular vesicles, exosomes function as an intercellular communication channel for different proteins and microRNAs (miRNAs) even for distant cells.[13] Recently, many studies have been conducted to find a relationship between exosomes secreted from specific cells and their impact on certain diseases. Such useful functions of exosomes have led scientists to utilize them for therapeutic purposes, one of them being HD.

MicroRNAs are small, non-coding strings of RNA with several functions that heavily impact gene expression which is why their specific role in genetic diseases and the overall human anatomy are being studied extensively. For example, one known function is how miRNAs regulate genes by creating bonds with the three untranslated regions (UTR) of messenger RNAs (mRNAs) to cause deregulation of the target gene's expression, essentially giving them the power to change the expression of multiple genes within a cell. This unique characteristic can be linked to potential positive or negative impacts the lack or abundance of specific miRNAs can have in a genetic disease.[13] Some recent studies have even proposed links between miRNAs and exosomes as there has been speculation around miRNA secretion being controlled by vesicular/exosomal-controlled mechanisms. Either way, exosomes do have a significant role within circulatory miRNA biology which is why exosomes have been used as transportation methods for miRNAs in some studies in order to facilitate gene exchange. Since miRNAs cannot move on their own, scientists have proposed that miRNAs are encapsulated and transported by exosomes when leaving or entering cells.[14]

Methods

A172 Glioblastoma Cell Culture:

A172 glioblastoma cells from homo sapiens brain cells were cultured in RPMI 1640 media supplemented with 10% fetal bovine serum (ATCC) and 1% penicillin and streptomycin. Media was exchanged every three days during cell maintenance. 0.25 % Trypsin-EDTA was used to detach the cells from culture flasks.

Neuronal Differentiation by Retinoic Acid Treatment:

A172 glioblastoma cells were differentiated to neuronal cells by adding retinoic acid. The 50 mM of stock solution of the retinoic acid was prepared. 0, 10, 30, 50 μM concentrations were used to differentiate the retinoic acid.

6-OHDA-derived HD Model:

The 100 mM of stock solution of 6-OHDA, a form of dopamine that has been used in previous Parkinson's studies to induce neuron damage, was prepared. The diluted concentrations of 6-OHDA were prepared and incubated with the A172 glioblastoma cells. IC50 was calculated using the Prism 7 program.

Exosome Isolation :

To isolate the exosomes from the A172 cell, the Total Exosome Isolation Reagent (Invitrogen) and exosome depleted fetal bovine serum (FBS) were used. After harvesting the cell culture media, the cell supernatant was centrifuged at 2000 x g for 30 minutes to remove cells and debris. Then the supernatant was moved to a new tube. The 500 μl of exosome isolation media was mixed with the 1 ml culture media. After the cell culture media and reagents were mixed well by vortexing, the samples were incubated at 4 ºC overnight. Then, the samples were centrifuged at 10,000 x g for 1 hour at 4 ºC. The supernatant was discarded and the pellet containing the exosomes was resuspended by PBS buffer.

miRNA Transfection:

0.5 x 106 cells were prepared in the 6 well culture plate. The miRNA and RNAimax transfection reagent (Invitrogen) were mixed together to 1:3 molar ratio to form a complex. Then the isolated exosomes were added to the mixture and incubated for 10 minutes. Then, the miRNA-RNAimax-exosome complex was added to the prepared cells.

Cell Viability Measurement:

PrestoBlue (Thermofisher) was used to measure the cell viability. After the cells were prepared in the 96 well plate, 10 μl of PrestoBlue reagent was added to each cell containing wells. Then, the cells were incubated in the CO_2 incubator at 37 ºC. After incubation, Epoch microplate reader was used to measure the absorbance at 570 nm wavelength.

Statistical Test:

All statistical tests were performed by the Prism 8 program. For all statistical analyses, a student t-test was used to calculate the p-value. The p-value lower than 0.05 was considered statistically significant.

Results and Discussion

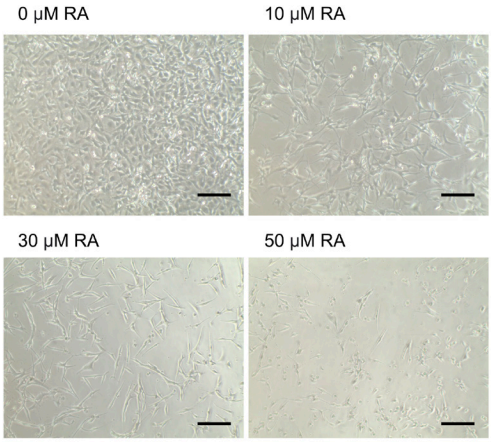


Figure 1: Microscopic images of A172 Glioblastoma cells treated by varying levels of retinoic acid (RA) for 24 hours. This result suggests that 30 μM concentration was the optimal concentration to differentiate A172 cells using RA. Scale bar = 100 μm.

To find the optimal concentration of RA treatment for neuronal differentiation, three different concentrations (10, 30, 50 μM) were treated on A172 cells as shown in Figure 1. After 24 hours of RA treatment, cell morphologies were observed. With increasing concentration of RA treatment, A172 cells elongated, exhibiting neuron-like morphology. However, at

DOI: 10.36838/v4i4.21

50 μM concentration, a large number of dead cells were observed, indicating possible toxicity of the substance. As a result of qualitative observation of A172 cells after RA treatment, 30 μM concentration was found to be the optimal concentration to induce differentiation, therefore this concentration will be used for the following experiments. Looking at the cell's morphology, the sample treated with 0 μM of retinoic acid has several small pieces with the occasional dark or white spot. The rest of the samples, treated with 10, 30, and 50 μM of retinoic acid, all have a generally round shape that grows in size as the concentration of retinoic acid increases, showing the progression of cell division. These samples seem to be epithelial with a long, stretched morphology. All cell samples are attached to the bottom layer of the petri dish and the image was taken from a bird's eye view.

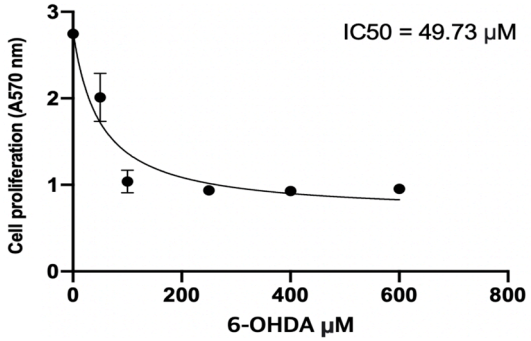


Figure 2: Cell viability of A172 glioblastoma cell under 6-OHDA treatment. IC50 value is indicated in the graph. This experiment result indicates which concentration of 6-OHDA is able to cause 50% of cell death in the A172 cells.

Varying concentrations of 6-OHDA were treated on A172 glioblastoma cells to observe their effects on cell viability by measuring 570 nm wavelength absorbance (Figure 2). Cell viability decreased exponentially upon 6-OHDA treatment. Increasing the concentration above 600 μM did not have any further effect on the proliferation rate. According to the best fit curve, corresponding IC50, which indicates the concentration needed to kill 50 % of the cells, concentration was found to be 49.73 μM. This concentration is used for the following experiments to induce HD-like conditions.

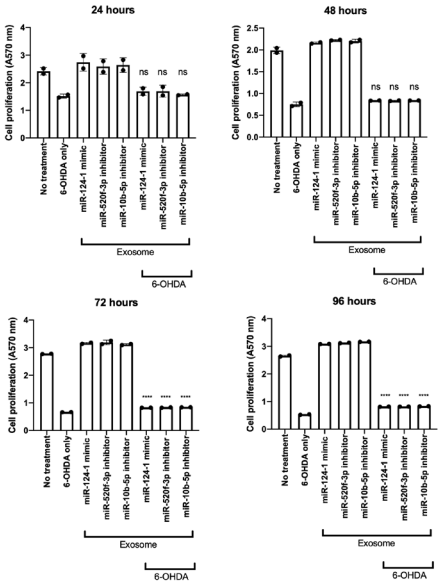


Figure 3: Effect of exosome-derived delivery of three types of miRNA on 6-OHDA-induced cell death during 24, 48, 72, 96 hours incubation. This result shows the long-term incubation of miRNAs and compares the results of manipulating the presence/absence of the miRNAs to 6-OHDA only and 6-OHDA + exosome + miRNA sample. 6-OHDA only sample was set as a control sample and an unpaired student t-test was performed to analyze the statistical significance. $p > 0.05$ (ns), $p < 0.001$ (****).

Effects of three types of miRNA delivered by exosome on A172 cell viability were analyzed over the period of 96 hours (Figure 3). 49.73 μM of 6-OHDA, as found in the previous section, was exposed to cells to mimic the condition of HD. 6-OHDA treatment consistently decreased cell viability with the progression of time, as the lowest value of absorbance was measured after 96 hours. For all three types of miRNA, miRNA and exosome combination seemed to increase the viability of cells for both control group and HD mimic group at all time points. The cell viability did not exhibit any significant differences between the three types of miRNAs after 24- and 48-hours incubation. There is no statistically significant difference in cell proliferation after 24 and 48 hours between 6-OHDA only samples and miRNA transfected samples. For the miRNA combinations to take effect, they must be treated for a minimum of 72 hours, as shown by the p-values derived above.

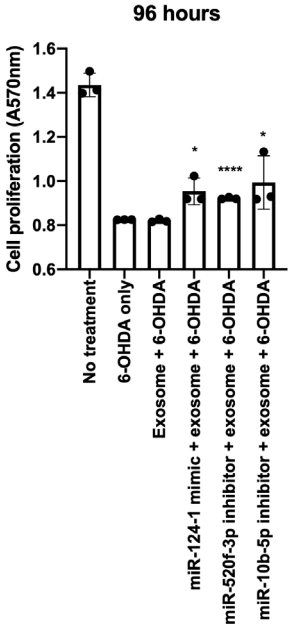


Figure 4: Effect of exosome and exosomal-miRNA combination on A172 cell viability in 6-OHDA-derived HD model after 96 hours. This figure shows the protective effect of miRNA on A172 cells, indicating that the protective effect of miRNA and exosome complexes in Figure 3 was not caused by the exosomes. Exosome + 6-OHDA sample was set as a control sample and an unpaired student t-test was performed to analyze the statistical significance. $p < 0.05$ (*), $p < 0.001$ (****)..

Data from Figure 4 shows there was a significant positive impact of the miRNA + exosome + reagent on cell proliferation after they were treated with 6-OHDA. However, the positive effect on cell proliferation may be induced either by exosomes with reagent or miRNA itself. To find out which factor derives the protective effect against 6-OHDA induced cell death, two conditions were tested with the addition of 6-OHDA:

1) exosomes with reagents 2) exosomes with reagents + miRNA.

Figure 4 shows that there is almost no difference in cell proliferation between cells treated only with 6-OHDA versus those treated with the exosome + reagent combination while there's a clear increase in cell proliferation in samples treated with the miRNA types as well showing that the impact of the last three bars is solely due to the addition of different miRNA types. Also, the error bars in the second and third bar graphs are almost zero, meaning this data is very reliable, again showing that the exosome + reagent combination had no positive impact on cell proliferation. On the other hand, the error bars for the fourth and last columns are quite significant but still show an increase in cell proliferation since the lower and upper bounds of all error bars still have a higher level of cell proliferation when compared to the second, 6-OHDA only bar. Therefore, the protective abilities of all three miRNA types against cells dying from 6-OHDA have been demonstrated.

Conclusion

Prior to the experiment, it was predicted that the insertion of certain miRNA types would increase proliferation in cells dying from HD, mimicked using 6-OHDA. Since the miRNA types were chosen based on pre-existing knowledge about miRNAs that are types impacted in the brains of HD patients, an experiment targeted towards ameliorating this shortcoming should have positively impacted cell growth that was already dying from HD. This hypothesis was supported as shown by Figures 3 and 4 as they clearly show that all three miRNAs significantly enhanced cell viability under 6-OHDA treatment.

While the change may seem trivial, considering that this experiment is exploring unknown topics with a lot of room to grow, this initial data is promising. Not only does it use 6-OHDA to mimic HD, but it also shows that miRNAs are effective in protecting cells from neuronal damage. This study is one of the first to explore how miRNAs can be used as a therapeutic method against HD and, unlike others that focus on one specific type, miR-124, this study explores the possibility of three additional types. One extension this research can take is to use combinations of the three types tested above. Since all three miRNAs have shown a positive impact on cell proliferation, their combination may further enhance the effect. If miRNAs prove to be an effective therapeutic in fighting HD, this disease might have its first cure. Granted, not every patient has the same circumstances so there cannot be just one solution, but by finding protective miRNAs and potentially using different combinations for each patient, this can be a potentially effective way to at least slow down disease progression.

In other studies that have used miRNAs and exosomes to discover more about HD, most have used cells from another species to transfect into a miRNA vector. For example, a previous study used genetic cloning technology to replicate their target miRNA into a plasmid to create miRNA vectors. They then transfected HEK 293 human cells with the miRNA vector and the resulting cells were put into a medium where scientists later chose cells that overexpressed the target miRNA.[15] The chosen cells were cultured into an exosome-free medium where the exosomes in the cell were isolated and the level of target miRNA left was quantified. To finish, the scientists injected the exosomes they harvested, that now had their target miRNA, into transgenic mice using a syringe.

This experiment, on the other hand, has differences that improve the efficacy and accuracy of the previously mentioned process. First, this experiment eliminates the need for vectors and cultured cells, allowing for the target and origin cell for exosome isolation to be from the same subject. Much like the idea behind organ transplants, it is much more preferable and beneficial if the original cell where an exosome was derived has a similar genetic makeup as the target cell as there is a significantly lower possibility of the target cell rejecting the exosome. However, this concept is not applied to the current, widespread method of miRNA delivery which is why this study proposes a method that theoretically would work. Therefore, this study will be able to explore the possibility of a method that can produce remarkably more accurate results thanks to the use of native cells.

As a whole, discovering effective miRNA types to recover and heal the neuron cells damaged from HD will be the first plausible cure/management method for the disease. While the specific method for injecting miRNAs into living human brains can be a future extension, this study shows the strong possibility of this treatment method panning out theoretically.

Acknowledgements

I would like to express my gratitude to my mentor, Lee Woo, as well as my parents who have supported me throughout this project on researching the protective abilities of miRNA types against Huntington's Disease.

References

1. McColgan, P.; Tabrizi, S. J., Huntington's disease: a clinical review. *European journal of neurology* **2018**, *25* (1), 24-34.
2. Ehrlich, M. E., Huntington's disease and the striatal medium spiny neuron: cell-autonomous and non-cell-autonomous mechanisms of disease. *Neurotherapeutics* **2012**, *9* (2), 270-284.
3. Duester, G., Retinoic acid synthesis and signaling during early or ganogenesis. *Cell* **2008**, *134* (6), 921-931.
4. Rochette-Egly, C., Retinoic acid signaling and mouse embryon ic stem cell differentiation: Cross talk between genomic and non-genomic effects of RA. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids* **2015**, *1851* (1), 66-75.
5. Zhang, J.; Gao, Y.; Yu, M.; Wu, H.; Ai, Z.; Wu, Y.; Liu, H.; Du, J.; Guo, Z.; Zhang, Y., Retinoic acid induces embryonic stem cell differentiation by altering both encoding RNA and microRNA ex pression. *PloS one* **2015**, *10* (7), e0132566.
6. McAllister, A., Neurotrophins and neuronal differentiation in the central nervous system. *Cellular and Molecular Life Sciences CMLS* **2001**, *58* (8), 1054-1060.
7. Bano, D.; Zanetti, F.; Mende, Y.; Nicotera, P., Neurodegenerative processes in Huntington's disease. *Cell death & disease* **2011**, *2* (11), e228-e228.
8. Cepeda, C.; Levine, M. S., Synaptic Dysfunction in Huntington's Disease: Lessons from Genetic Animal Models. *The Neuroscientist* **2020**, 1073858420972662.
9. André, V. M.; Cepeda, C.; Levine, M. S., Dopamine and glutamate in Huntington's disease: A balancing act. *CNS neuroscience & therapeutics* **2010,** *16* (3), 163-178.

DOI: 10.36838/v4i4.21

10.Simola, N.; Morelli, M.; Carta, A. R., The 6-hydroxydopamine model of Parkinson's disease. *Neurotoxicity research* **2007**, *11* (3), 151-167.
11.Rea, S.; Della-Morte, D.; Pacifici, F.; Capuani, B.; Pastore, D.; Coppola, A.; Arriga, R.; Andreadi, A.; Donadel, G.; Di Daniele, N., Insulin and Exendin-4 Reduced Mutated Huntingtin Accu mulation in Neuronal Cells. *Frontiers in pharmacology* **2020**, *11*, 779.
12.Zhang, Y.; Liu, Y.; Liu, H.; Tang, W. H., Exosomes: biogenesis, biologic function and clinical potential. *Cell & bioscience* **2019**, *9* (1), 1-18.
13.Zhang, J.; Li, S.; Li, L.; Li, M.; Guo, C.; Yao, J.; Mi, S., Exosome and exosomal microRNA: trafficking, sorting, and function. *Genomics, proteomics & bioinformatics* **2015**, *13* (1), 17-24.
14.Zhang, D.; Lee, H.; Zhu, Z.; Minhas, J. K.; Jin, Y., Enrichment of selective miRNAs in exosomes and delivery of exosomal miRNAs in vitro and in vivo. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **2017**, *312* (1), L110-L121.
15.Fan, J.; Feng, Y.; Zhang, R.; Zhang, W.; Shu, Y.; Zeng, Z.; Huang, S.; Zhang, L.; Huang, B.; Wu, D., A simplified system for the ef fective expression and delivery of functional mature microRNAs in mammalian cells. *Cancer gene therapy* **2020**, *27* (6), 424-437.

Author

Yejin Kong is a student at Seoul Foreign School in South Korea, expected to graduate in 2023. She shows interest in pursuing Molecular Biology and/or Biochemistry in her future studies, potentially going into the field of medicine. She has completed courses in Microbiology and Diseases at UC Berkeley and Brown Summer Camps and has participated in CHLA's (Children's Hospital in Los Angeles) Career Exploration Program as well as the FDA's (Food and Drug Association) Oncology Center of Excellence Summer Scholars Program. She has also received the "Excellent Youth Scholars Award" from KSCY (Korea Scholar's Conference for Youth) for scientific research and earned a Silver Medal in Genius Olympiad.

Marine Biology Research at Bahamas

Unique and exclusive partnership with the Gerace Research Center (GRC) in San Salvador, Bahamas to offer marine biology research opportunities for high school teachers and students.

- Terra has exclusive rights to offer the program to high school teachers and students around world.
- All trips entail extensive snorkeling in Bahamian reefs as well as other scientific and cultural activities.
- Terra will schedule the program with GRC and book the flights from US to the GRC site.
- Fees include travel within the US to Island, lodging, meals, and hotels for transfers, and courses.
- For more information, please visit terraed.org/bahamas.html

Terra is a N.Y. based 501.c.3 non-profit organization
dedicated for improving K-16 education

IJHSR International Journal of High School Research

is a publication of

terra science and education

N.Y. based 501.c.3 non-profit organization
dedicated for improving K-16 education