IJHSR
ijhighschoolresearch.org
■ RESEARCH ARTICLE

Machine Learning for miRNA Biomarker Identification in Ovarian Cancer

Laura Chen[1], Sophia Kyveryga[1], Hongyang Gao[2]

1. Ames High School, 1801 Ridgewood Ave, Ames, Iowa, 50010, USA; lauraychen03@gmail.com
2. Iowa State University, Ames, Iowa, 50011, USA

ABSTRACT: Ovarian cancer is the 5th-deadliest cancer for women worldwide. 80% of ovarian cancer cases are caught in late Stages III and IV when the metastasis of the malignant tumor has already occurred. This study leverages machine learning algorithms to identify up- or down-regulated miRNAs using the genetic data of ovarian cancer patients from The Cancer Genome Atlas (TCGA). Based on a comparative framework of six machine learning models validating its applicability, Principal Component Analysis (PCA) was adopted as the primary machine learning model for both biomarker identification and dimension reduction of the genomic dataset. Through PCA's Explained Variance Ratio (EVR) and miRNA loadings, a dataset of miRNA biomarkers for ovarian cancer was obtained. Among them, the top 10 candidate miRNA biomarkers for ovarian cancer were corroborated by existing literature, which validates the method used in this study. Additionally, 10 candidate miRNAs were proposed that had not been previously identified as biomarkers in the existing literature, which could be a potential path for exploration as biomarkers for ovarian cancer. This study proves the applicability of identifying miRNA biomarkers for ovarian cancer using machine learning and lays a foundation for future research.

KEYWORDS: Computational Biology and Bioinformatics, Geonomics, Machine Learning, Ovarian Cancer, miRNA Biomarker.

Introduction

Cancer is a global health crisis, accounting for about 10 million deaths each year.[1] Regular cancer screening is a core part of maintaining good health. Notably, ovarian cancer, the 5th deadliest cancer for women, has no regular screening option,[2] and its symptoms are vague and often mistaken for less-serious conditions. As shown in Figure 1, ovarian cancer caught in Stage I (only in the ovaries) can be cured in up to 90% of patients. In contrast, from Stage III onwards, the 5-year survival rate of ovarian cancer drops substantially. However, only 20% of ovarian cancer cases are caught in Stages I and II.[3] As a result, the 5-year survival rate of ovarian cancer is significantly lower than that of other types of cancer, as illustrated in Figure 2.

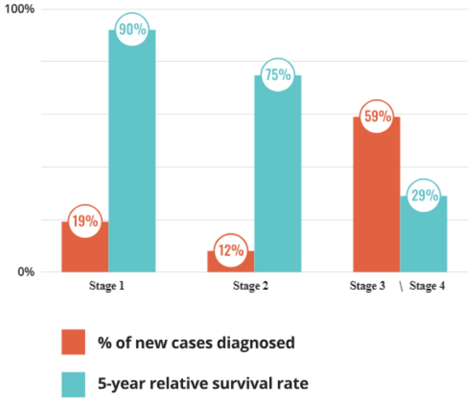


Figure 1: Proportion of new cases diagnosed and 5-year-survival rates by ovarian cancer stages.[4] This graph shows that ovarian cancer caught in Stage I (only in the ovaries) can be cured in up to 90% of patients. In contrast, from Stage III onwards, the 5-year survival rate of ovarian cancer drops substantially.

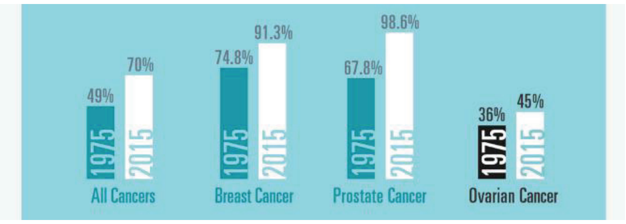


Figure 2: 5-year-survival rates of ovarian cancer in 2015 vs. 1975 compared to other cancers.[5] This figure shows that the 5-year survival rate of ovarian cancer is significantly lower than that of other types of cancer.

MiRNAs (microRNAs) are single-stranded and non-coding RNA molecules that regulate gene expression by binding to messenger RNAs. miRNAs circulate in the body and are found in bodily fluids such as blood, urine, and saliva. In the last two decades, miRNAs have shown potential as a biomarker for cancer, especially in detecting cancer in the body. MiRNAs' quantifiability, resistance to harsh storage conditions, high stability in plasma and serum, and detectability in blood make them an ideal choice as biomarkers. Additionally, advances are being made in the miniaturization of devices that detect circulating biomarkers in the body, making it less invasive, more rapid, and more realistic to use miRNA as a biomarker.[6] The close relationship between miRNA expression and the specific biological pathways they control enables miRNAs to play a substantial role in early screening and prognosis, targeted therapy, and drug resistance tendencies in ovarian cancer.[7] This is because cancer mechanisms, such as epigenetic changes, mutations, chromosomal abnormalities, and defects in the miRNA biogenesis pathway, can silence mRNA expression. Therefore, miRNA shows promise as a biomarker for cancer, since it can

DOI: 10.36838/IJHSR86.37

be easily detected in the blood and tissue of cancer-positive patients, especially for ovarian cancer.

Artificial Intelligence (AI) using machine learning (ML) has become a powerful tool for identifying patterns in high-dimensional datasets, such as genomic data. Recent studies have found ML applications to have a 98.2% accuracy in identifying cancers using genetic data.[8] In the context of ovarian cancer, which involves a dysregulation of microRNA expression that influences carcinogenesis, invasion, and metastasis, ML offers a promising way to identify miRNA biomarkers. Because of miRNAs' role in regulating gene expression and serving as either tumor suppressors or oncogenes, using their differential expression between cancer patients and healthy individuals can significantly improve early diagnosis. However, the large dimensionality of genomic data serves as a challenge for many classification models. Principal Component Analysis (PCA) can address this issue by serving as a dimension-reduction technique while preserving variance. Several studies have applied PCA to genomic datasets for cancer classification, including Hamidi *et al.*[9] and Kartikasari *et al.*[10] These studies confirm PCA's relevance in this facet. Our approach differs by introducing a comparative ML framework and leveraging the EVR-based biomarker ranking. Therefore, the paper aims to: (1) evaluate the effectiveness of PCA for the genomic dataset using six ML models; and (2) identify the top 30 miRNA biomarkers and validate them with existing literature. This study not only reinforces the use of PCA in handling genomic data reduction but also highlights its potential in identifying notable biomarkers.

Methodology

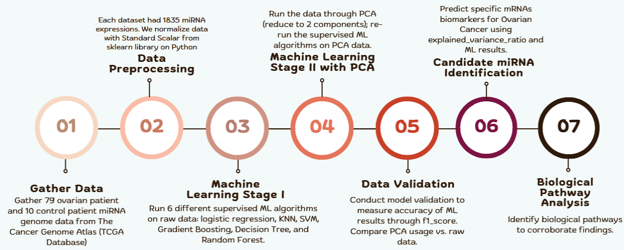


Figure 3: Flowchart of the method. This diagram shows the steps of the methodology, including Gather Data, Data Processing, Machine Learning Stage I, Machine Learning Stage II with PCA, Data Validation, Candidate miRNA Identification, and Biological Pathway Analysis.

Figure 3 shows the flowchart of the method, with the following steps:

Gather and Process Data:

The Cancer Genome Atlas (TCGA) provided miRNA expression data for ovarian cancer patients and healthy controls. The dataset included 79 ovarian cancer and 10 control samples, each with 1,835 miRNA expression values. The label column is the binary column (0 for control and 1 for cancer), and the feature columns span the other 1,800+ columns.

Principal Component Analysis (PCA):

The flowchart of the method is shown in Figure 4, detailing the integration of PCA in the workflow. PCA was applied to reduce the entire dataset to two and three principal components, respectively. PCA was used as the dimension reduction technique, and the EVR loadings assigned a numerical value of importance to each miRNA, capturing the data variance. The choice of two and three principal components was guided by the amount of variance captured. The first three principal components captured 0.14, 0.07, and 0.04 of the data variation, respectively. Based on the curve of the scree plot (Figure 11 in the Appendix) and a clear divide between the control and ovarian cancer patients on the graphs, we selected up to 3 components. We compared the performance of the ML models on lower-dimensional data versus the original dataset to ensure the applicability of PCA to the given dataset.

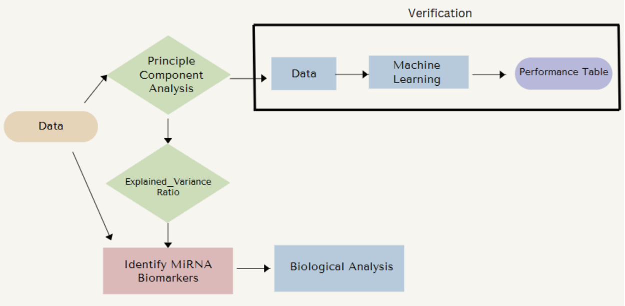


Figure 4: Flowchart of PCA application. This figure details the integration of PCA in the workflow.

Machine Learning Performance Table:

Six ML models, namely Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Gradient Boosting, Decision Tree, and Random Forest, were used in the comparative framework. These models were chosen for their effectiveness in classification tasks that use high-dimensional data, such as genomic data. The models were performed on the 2-dimensional PCA data, the 3-dimensional PCA data, and the raw data. Each dataset was stratified and split into a training and a testing dataset.

The accuracy of each ML model's prediction was measured on the datasets using the F1-score for the testing set. The F1-score is a harmonic mean between precision and recall, which is particularly useful for imbalanced datasets such as TCGA, where the number of cancer samples (n=79) exceeds the number of control samples (n=10). F1-score has been used in previous genomic and miRNA-based cancer classification tasks, such as those in Hamidi *et al.*[9] and Bhardwaj *et al.*[11] This step identified the applicability of PCA to this dataset by comparing the accuracy of the results between the simplified and raw datasets.

Candidate miRNA Identification:

After verifying the performance table to ensure PCA applicability, EVR was applied to evaluate the most significant miRNAs. To identify and rank the significant miRNAs for ovarian cancer classification, we used the loadings of the individual miRNAs on the top principal components that captured the most variance. This approach ranked miRNAs by relative importance, focusing on the miRNAs that differed in expression between the cancer and control groups. Upregulation and downregulation were calculated by subtracting the mean of the two groups. Open databases miRWalk and miRDB were then used to validate the roles of the significant miRNAs and identify the target genes of the selected miRNAs.

■ Data Analysis

Data Pre-Processing:

By applying StandardScaler from the sci-kit learn library, we constructed a new data frame that allowed for more effective use of PCA. Each column was standardized so that the mean was 0 and the standard deviation was 1, allowing for more balanced consideration of each miRNA's significance.

Principal Component Analysis (PCA):

PCA was used as a dimensionality reduction technique to identify the most important features (principal components) that capture the most variance while discarding less important data. It effectively reduced the dataset dimensions into two and three components, respectively. After generating the coordinates and the graph of the 2-component PCA in Figure 5 (0 for control and 1 for cancer), we generated a 3-component coordinate and graph to serve as a point of comparison, as shown in Figure 6. Table 1 shows the variance explained and cumulative variance per PC from 1 to 3, where each PC represents a new axis that captures the most variance in the miRNA data. Tables 2 and 3 show the snippets of the dataset coordinates for the 2-component and 3-component PCA reductions, respectively.

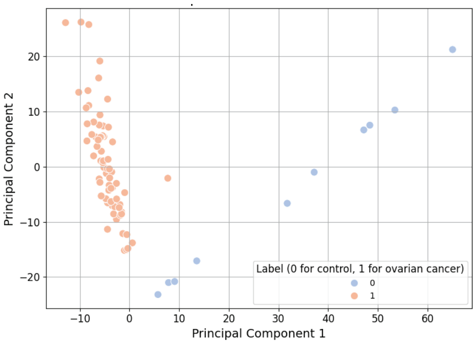


Figure 5: 2-Component PCA scatter plot. This figure shows the coordinates and the graph of the 2-component PCA (0 for control and 1 for cancer).

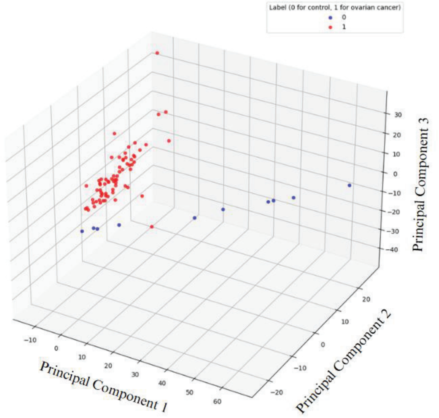


Figure 6: 3-Component PCA scatter plot. This figure shows the 3-component coordinates and graph.

Table 1: Variance explained and cumulative variance per PC from 1 to 3. Each PC represents a new axis that captures the most variance in the miRNA data.

Principal Component	Variance Explained	Cumulative Variance
PC1	0.1425	0.1425
PC2	0.0750	0.2175
PC3	0.0452	0.2627

Table 2: 2-Component PCA of miRNA data coordinates. This table shows a snippet of the dataset coordinates for the 2-component PCA reduction.

PC1	PC2	Label
-6.76393	5.38631	1
-4.67066	-1.21967	1
-5.77191	1.062171	1
-2.61922	-9.53806	1
-10.2212	13.49393	1
-8.204	11.10435	1
0.580015	-13.8038	1
-1.01663	-15.1722	1
-8.56846	4.67346	1
-3.4423	-3.81268	1
-5.16903	5.456385	1
-5.4201	5.614891	1

Table 3: 3-Component PCA of miRNA data coordinates. This table shows a snippet of the dataset coordinates for the 3-component PCA reduction.

PC1	PC2	PC3	Label
-6.76197	5.386978	2.177069	1
-4.66926	-1.21935	11.13995	1
-5.77033	1.062688	-3.13357	1
-2.61831	-9.53829	-1.91555	1
-10.2194	13.49576	-5.58639	1
-8.20174	11.1053	-1.81371	1
0.581108	-13.8048	3.258688	1
-1.01593	-15.1729	1.115581	1
-8.56699	4.674769	-6.04256	1
-3.44076	-3.81272	1.664314	1
-5.16682	5.456724	-11.5222	1
-5.41796	5.61539	0.563312	1

Machine Learning Performance Table:

Figure 7 compares the accuracy of six supervised ML models on 3 datasets (2D PCA, 3D PCA, and raw datasets). The model performance was highest on the raw dataset and the 3-component PCA, while the 2-component PCA was suboptimal in the gradient boosting and decision tree model predictions. The data were split into a training and testing group, with 80% (71 samples) of the data allocated to the training set and 20% (18 samples) to the testing set. Overall, the high accuracy in predicting the testing group for all datasets signified the applicability of PCA for dataset reduction and variance capturing, validating the results from the EVR loadings.

DOI: 10.36838/IJHSR86.37

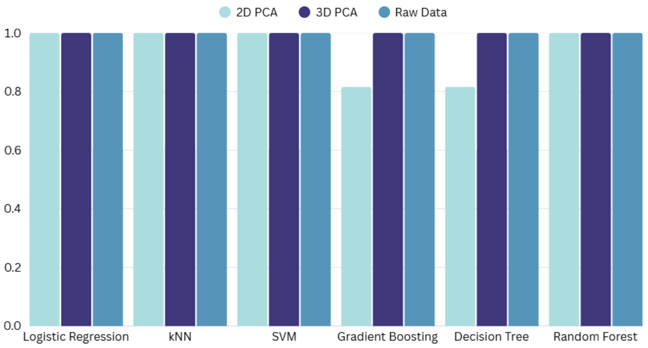


Figure 7: Comparison table of PCA and non-PCA data for six ML models. This figure compares the accuracy of six supervised ML models on 3 datasets (2D PCA, 3D PCA, and raw datasets).

Candidate miRNA Identification:

Using the EVR loadings, we created a list of miRNAs ranked by the highest loading ratio (most significant) to the lowest ratio (least significant), as shown in the Importance column of Table 4. We classified the miRNA as upregulated or downregulated by subtracting the mean expression level in the control group from the expression level in the ovarian cancer group. Figure 8 was generated to identify the trend in upregulation and downregulation in the top 30 miRNAs. Table 4 shows a snippet of the dataset after experimentation.

Table 4: Ranked miRNAs based on EVR loadings & expression classification. It shows a list of miRNAs ranked by the highest loading ratio (most significant) to the lowest ratio (least significant) in the Importance column and a snippet of the dataset after experimentation.

miRNA	Importance	Mean_0	Mean_1	Difference (1-0)	Regulation
hsa-mir-3158-2	0.02325936	-0.998708897	0.14474042	1.143449317	upregulated
hsa-mir-6862-1	0.023251593	-0.713832595	0.103453999	0.817286594	upregulated
hsa-mir-6770-1	0.023165822	-0.866469536	0.125575295	0.992044831	upregulated
hsa-mir-455	0.023136819	-1.162737852	0.168512732	1.331250585	upregulated
hsa-mir-153-1	0.023025013	-0.781940547	0.113324717	0.895265263	upregulated
hsa-mir-125a	0.022976412	-1.408040529	0.204063845	1.612104374	upregulated
hsa-mir-6862-2	0.022941704	-0.878782361	0.127359763	1.006142124	upregulated
hsa-mir-191	0.02282751	-1.311693819	0.190100554	1.501794373	upregulated
hsa-mir-424	0.022806606	-0.783537198	0.113556116	0.897093313	upregulated
hsa-mir-1275	0.022702766	0.68906242	-0.099864119	-0.788926539	downregulated
hsa-mir-423	0.022657594	-1.384881262	0.200707429	1.585588691	upregulated
hsa-mir-6886	0.022644081	-0.68877402	0.099822322	0.788596341	upregulated

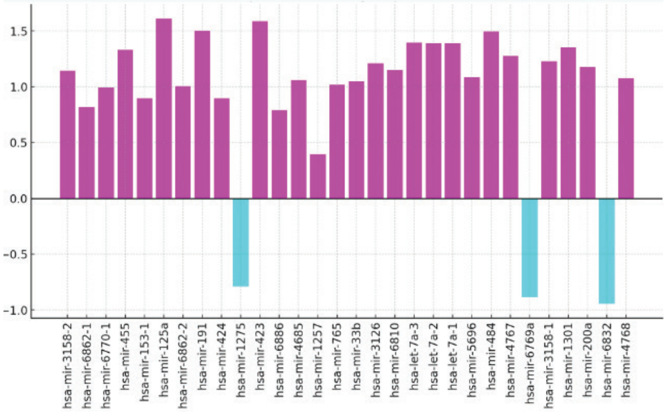


Figure 8: Regulation classification of the top 30 miRNAs. This figure identifies the trend in upregulation and downregulation in the top 30 miRNAs.

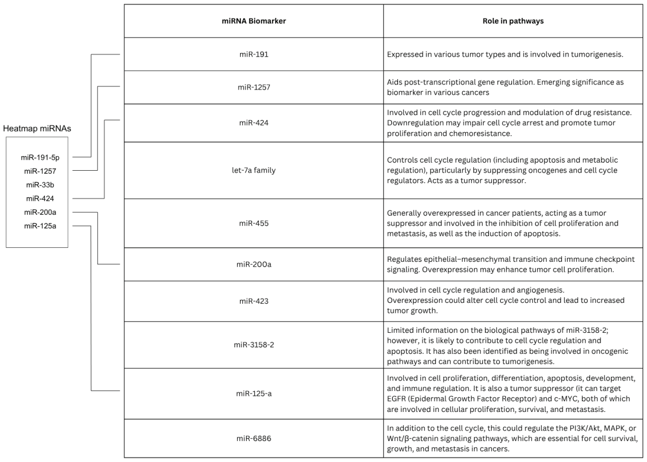
Biological Pathway Analysis:

We focused our analysis on the top 30 miRNAs out of the total miRNAs identified through experimentation. We used miRDB to generate a heatmap of the known pathways of 5 candidate miRNAs identified in similar cancer pathways using the KEGG database and miRPathDB, as shown in Figure 9. Our findings corroborated the database for the notable miRNAs, placing them in the top 30 most significant in differential expression for ovarian cancer.

Figure 9: Heatmap of known pathways of candidate miRNAs using the KEGG database.[12] This figure shows a heatmap of the known pathways of 5 candidate miRNAs identified in similar cancer pathways using the KEGG database and miRPathDB.

We utilized open-source biological pathways, such as miRDB and miRPathDB, to identify the biological pathways of the candidate miRNAs and provide a preliminary explanation for the importance of their expression levels in ovarian cancer and control patients, as summarized in Table 5.

Table 5: Candidate miRNA and pathway analysis for notable candidate miRNAs as a linking chart to Figure 9. This table identifies the biological pathways of the candidate miRNAs and provides a preliminary explanation for the importance of their expression levels in ovarian cancer and control patients.



Heatmap miRNAs

miR-191-5p
miR-1257
miR-33b
miR-424
miR-200a
miR-125a

miRNA Biomarker	Role in pathways
miR-191	Expressed in various tumor types and is involved in tumorigenesis.
miR-1257	Aids post-transcriptional gene regulation. Emerging significance as biomarker in various cancers
miR-424	Involved in cell cycle progression and modulation of drug resistance. Downregulation may impair cell cycle arrest and promote tumor proliferation and chemoresistance.
let-7a family	Controls cell cycle regulation (including apoptosis and metabolic regulation), particularly by suppressing oncogenes and cell cycle regulators. Acts as a tumor suppressor.
miR-455	Generally overexpressed in cancer patients, acting as a tumor suppressor and involved in the inhibition of cell proliferation and metastasis, as well as the induction of apoptosis.
miR-200a	Regulates epithelial–mesenchymal transition and immune checkpoint signaling. Overexpression may enhance tumor cell proliferation.
miR-423	Involved in cell cycle regulation and angiogenesis. Overexpression could alter cell cycle control and lead to increased tumor growth.
miR-3158-2	Limited information on the biological pathways of miR-3158-2; however, it is likely to contribute to cell cycle regulation and apoptosis. It has also been identified as being involved in oncogenic pathways and can contribute to tumorigenesis.
miR-125-a	Involved in cell proliferation, differentiation, apoptosis, development, and immune regulation. It is also a tumor suppressor (it can target EGFR (Epidermal Growth Factor Receptor) and c-MYC, both of which are involved in cellular proliferation, survival, and metastasis.
miR-6886	In addition to the cell cycle, this could regulate the PI3K/Akt, MAPK, or Wnt/β-catenin signaling pathways, which are essential for cell survival, growth, and metastasis in cancers.

The use of miRNA as a biomarker shows promise as an indicator of early-stage ovarian cancer, increasing a patient's chance of survival substantially. Based on our findings, we identified miRNAs that have been corroborated by previous literature, as well as other notable miRNAs not previously reported, which may serve as effective therapeutic targets or a means to combat drug-resistant tendencies in future ovarian cancer treatments.[13]

DOI: 10.36838/IJHSR86.37

We analyzed the relationship between the top 30 miRNAs by generating a heatmap to identify patterns in expression levels across the patients, as shown in Figure 10. We found that the let-7a family was heavily connected in terms of differential expression levels, as well as hsa-mir-3158-1 and hsa-mir-3158-2.

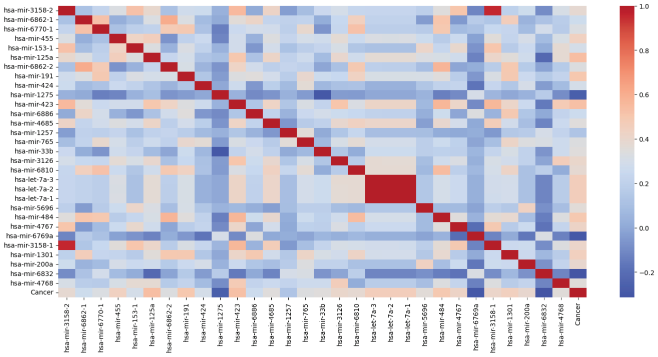


Figure 10: Correlation between expression levels for the top 30 miRNAs. This figure shows a heatmap to identify patterns in expression levels across the patients.

Results and Discussion

We identified the top 10 candidate miRNA biomarkers for ovarian cancer, corroborated by the current literature, and ranked them in accordance with our list in Table 6. This validated the results of our experimental method, as it achieved similar results to those in previous literature that had used physical experimentation.

Table 6: Candidate miRNAs corroborated by current literature with rankings. This table shows the top 10 candidate miRNA biomarkers for ovarian cancer corroborated by the current literature.

Ranking	miRNA
10	hsa-miR-424
11	hsa-miR-1275
20	hsa-let-7a-3
21	hsa-let-7a-2
22	hsa-let-7a-1
24	hsa-miR-484
29	hsa-miR-200a
44	hsa-miR-221
57	hsa-miR-502
61	hsa-miR-429

Notably, Pal *et al.*[14] corroborated the following miRNAs: let-7a-3, let-7a-2, let-7a-1, miR-200a, miR-221, miR-484, and miR-429. Zhao *et al.*[13] corroborated the following miRNAs: let-7a-3, let-7a-2, let-7a-1, miR-424, miR-502, miR-484, miR-429, miR-200a, and miR-221.

Additionally, we noted our top 10 candidate miRNAs that had not been identified by existing literature, as shown in Table 7. We introduce them as a potential path for exploration as a biomarker for ovarian cancer.

Table 7: Candidate miRNAs that have not been introduced in the current literature with rankings. This table shows the top 10 candidate miRNAs that had not been identified by existing literature.

Ranking	miRNA
1	hsa-miR-3158-2
2	hsa-miR-6862-2
3	hsa-miR-6862-1
4	hsa-miR-191
5	hsa-miR-6770-1
6	hsa-miR-423
7	hsa-miR-455
8	hsa-miR-6886
9	hsa-miR-125a
12	hsa-miR-4685

Conclusion

Early detection of cancers can significantly increase the survival rate. This paper presents a study aimed at identifying miRNA biomarkers for ovarian cancer using machine learning. Based on the study, the following conclusions can be drawn:

(1) The unique patient dataset from the TCGA serves as a reliable source that can strengthen existing research on miRNA biomarkers.

(2) Based on a comparative framework of six machine learning models validating its applicability, PCA is shown to be an efficient and applicable way to identify significant miRNA expression trends between cancer patient and healthy control genomes, and holds potential in future biomarker discovery for cancers, notably ovarian cancer.

(3) Using PCA's EVR and miRNA loadings, a dataset of miRNA biomarkers for ovarian cancer was obtained. Among them, the top 10 candidate miRNA biomarkers were corroborated by existing literature, which validates the method used in this study. Additionally, several notable miRNAs were identified that have not been reported in the existing literature, which can be used for the earlier identification of and targeted therapies for ovarian cancer.

This study proves the applicability of identifying miRNA biomarkers for ovarian cancer using machine learning and lays a foundation for future research, including:

(1) Diversify patient data to corroborate findings: Branch out from The Cancer Genome Atlas Data and use diverse ethnic backgrounds.

(2) Use real-world experimentation to validate results, such as quantitative polymerase chain reaction (qPCR), to corroborate results in key miRNA biomarkers.

One limitation of this study is that the TCGA dataset is derived from solid tissue samples. To increase practicality for early detection in clinical settings, future work will focus on using blood-based assays or liquid biopsies to extend the PCA framework results. Applying machine learning to these non-invasive samples will strengthen the applicability of our findings.

Based on the results, the next step to further strengthen these findings will be validating candidate miRNAs with independent datasets beyond TCGA. Publicly available resources, such as the Gene Expression Omnibus (GEO), ArrayExpress, and the International Cancer Genome Consortium (ICGC), provide diverse patient backgrounds, which can enhance the generalizability of the results.

Appendix

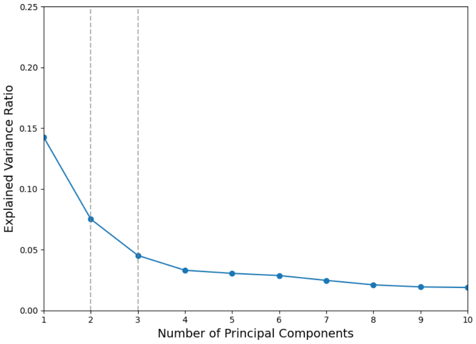


Figure 11: Scree plot of explained variance ratio per principal component. This figure justifies the use of 3-D PCA by showing the decrease in variance explained per additional PC after 3 components.

Acknowledgments

The student authors would like to express sincere gratitude to Dr. Gao for his insight and guidance throughout this research project. We also want to acknowledge Dr. Ronnie Li for his support in the project ideation phase. Additionally, we want to thank the judges at the 2025 SSTFI (State Science and Technology Fair of Iowa) for their feedback and nomination in the Biomedical and Health Sciences category.

References

1. "Cancer Today." *Global Cancer Observatory*. https://gco.iarc.fr/today/en. Accessed 12 December 2024.
2. UCLA Health. "How is ovarian cancer detected?" *UCLA Health*. 31 October 2022, https://www.uclahealth.org/news/article/how-is-ovarian-cancer-detected. Accessed 12 December 2024.
3. Elias, K. M., Guo, J., Bast, R. C. Jr. "Early Detection of Ovarian Cancer." *Hematol Oncol Clin North Am*. 2018, *32*(6), 903-914. https://doi.org/10.1016/j.hoc.2018.07.003. PMID: 30390764; PMCID: PMC6376972.
4. Ovarian Cancer Research Foundation. "About Ovarian Cancer." *Ovarian Cancer Research Foundation*. 2023. www.ocrf.com.au/ovarian-cancer/facts-about-ovarian-cancer.
5. Ovarian Cancer Research Foundation. "Why Is Ovarian Cancer so Tricky to Treat?" *Ovarian Cancer Research Foundation*. 29 Nov. 2021. www.ocrf.com.au/news/68/why-is-ovarian-cancer-so-tricky-to-treat.
6. Gayosso-Gómez, L. V., Ortiz-Quintero, B. "Circulating MicroRNAs in Blood and Other Body Fluids as Biomarkers for Diagnosis, Prognosis, and Therapy Response in Lung Cancer." *Diagnostics (Basel)*. 2021, *11*(3), 421. https://pmc.ncbi.nlm.nih.gov/articles/PMC7999833/.
7. Chakrabortty, A., Patton, D. J., Smith, B. F., Agarwal, P. "miRNAs: Potential as Biomarkers and Therapeutic Targets for Cancer." *Genes (Basel)*. 2023, *14*(7), 1375. doi: 10.3390/genes14071375. PMID: 37510280; PMCID: PMC10378777.
8. Newsham, I. *et al*. "Early Detection and Diagnosis of Cancer with Interpretable Machine Learning to Uncover Cancer-Specific DNA Methylation Patterns." *Biology Methods and Protocols*. 2024, *9*(1). https://doi.org/10.1093/biomethods/bpae028.
9. Hamidi, F., *et al*. "Identifying potential circulating miRNA biomarkers for the diagnosis and prediction of ovarian cancer using a machine-learning approach: application of Boruta." *Frontiers in Digital Health*. 2023. *5*. https://doi.org/10.3389/fdgth.2023.1187578.
10. Kartikasari, A. E. R., *et al*. "Circulating microRNAs as Diagnostic Biomarkers to Detect Specific Stages of Ovarian Cancer: A Comprehensive Meta-Analysis." *Cancers*. 2024. *16*(24), 4190. https://doi.org/10.3390/cancers16244190.
11. Bhardwaj, A., *et al*. "Tree-Based and Machine Learning Algorithm Analysis for Breast Cancer Classification." *Computational Intelligence and Neuroscience*. 2022. *2022*, 1-6. https://doi.org/10.1155/2022/6715406.
12. Backes, C., Kehl, T. "MiRPathDB." Mpd.bioinf.uni-Sb.de, mpd.bioinf.uni-sb.de/heatmap_calculator.html?organism=hsa. Accessed 31 March 2025.
13. Zhao, L. *et al*. "The Role of MiRNA in Ovarian Cancer: An Overview." *Reproductive Sciences*. 2022, *29*(10), 2760–2767. https://doi.org/10.1007/s43032-021-00717-w.
14. Pal, M. K. *et al*. "MicroRNA: A New and Promising Potential Biomarker for Diagnosis and Prognosis of Ovarian Cancer." *Cancer Biology & Medicine*. 2015. *12*(4), 328–341. www.ncbi.nlm.nih.gov/pmc/articles/PMC4706521/, https://doi.org/10.7497/j.issn.2095-3941.2015.0024.

Authors

Laura Chen is a Senior at Ames High School. Her interest in the interdisciplinary nature of computing has brought her to bioinformatics. In her future studies, she intends to delve into further research on ML and its applications in data analytics.

Sophia Kyveryga is a Senior at Ames High School. Her long interest in mathematics and growing interest in biology have led her to the field of bioinformatics, with future plans to further explore paths such as biochemical engineering.

Dr. Hongyang Gao is a faculty member in the Department of Computer Science, Iowa State University. He received his Ph.D. degree from Texas A&M University. His research interests include AI, ML, Deep Learning, and Computational Biology.