IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Neuroflux: An Interpretable Multimodal U-Net Model for Glioblastoma Multiforme Segmentation Through Grad-CAM

Eve Ang Yee Leng[1], Darius Mihai Oltei[2], Pranav Vedula[3], Shriya Viswanathan[4]

1) Saint Joseph's Institution, 38 Malcolm Road, Singapore, Singapore, 308274, Singapore; eveangyeeleng@gmail.com
2) "Octavian Goga" National College, Bastionului Street, Sibiu, Sibiu, 550177, Romania
3) Colonial Forge High School, 550 Courthouse Rd, Stafford, VA, 22554, USA
4) Plano Senior High School, 2200 Independence Pkwy, Plano, TX, 75075, USA

ABSTRACT: Glioblastoma multiforme (GBM) is one of the most aggressive malignant brain tumors, with early intervention through diagnosis and treatment being critical to improving the low median overall survival rate of 15 months. Manual anatomical segmentation is problematic because it is prone to inconsistencies from interobserver variation. Current models are highly accurate but lack computational efficiency and multimodality, which limits the clinical applicability of such segmentation tools in hospital environments. Neuroflux is a multimodal U-Net-based deep learning model for GBM segmentation using the BraTS 2020 dataset. Adapted from a baseline model, we enhanced the preprocessing of raw data through MinMax scaling and applied our custom hybrid loss strategy to maintain high levels of accuracy while minimizing false negatives. With further improvements in regularization techniques, Neuroflux was also fine-tuned to incorporate targeted modifications and optimized dropout rates, which improved the training strategy. Overall, an accuracy of 99.01%, precision of 99.11%, sensitivity of 98.85%, and specificity of 99.7% was achieved while significantly improving clinician interpretability. Our high-precision and efficient modified architecture highlights the potential for new lightweight and interpretable models to enhance diagnostic reliability and treatment planning.

KEYWORDS: Computational Biology and Bioinformatics, Computational Neuroscience, Neuro-oncological Imaging, Brain Tumor Segmentation, Glioblastoma Multiforme.

Introduction

Background information:

GBM is a primary brain neoplasm comprising genetically and phenotypically heterogeneous groups of tumors.1 Being located in the hemispheres, or subtentorially in the brainstem and cerebellum, it is one of the most aggressive malignancies, with a very low median overall survival rate of only 15 months.[1,2] Due to its high complexity and invasive potential, predicting patient survival and response to treatments through tumor segmentation is critical for treatment decisions in GBM. As the most aggressive diffuse glioma of the astrocyte lineage, GBM remains an incurable tumor with a "wildtype" isocitrate dehydrogenase (IDH) status.[2,3] Traditional anatomical segmentation methods rely on manual pathologists' review, which can be very inconsistent due to interobserver variation.[3] This becomes highly problematic as accurate tumor segmentation from Magnetic Resonance Imaging (MRI) scans is crucial for diagnostic outcomes and treatment planning. Furthermore, these approaches often lack the robustness and scalability needed for precise and automated tumor classification. With the rise of machine learning in clinical radiology, convolutional neural networks (CNN) have been increasingly utilized in glioblastoma.[4-6] Artificial Intelligence (AI)-driven segmentation models have been evaluated as a promising alternative to manual segmentation.

Current limitations:

One of the core ethical concerns in utilizing AI in medical diagnosis is the lack of transparency in the model's processes. Also known as a "black box," there is a severe lapse in the lack of readability to medical professionals when interpreting the rationale behind a model's decision on the segmented tumor. Explainable AI (XAI) has been hailed in recent years as a solution to this "black box" due to its high application, particularly in oncology. Neuro-oncology is an especially critical subcategory for XAI due to the high stakes of patient diagnosis, which is a key factor in early intervention. Gradient-weighted Class Activation Mapping (Grad-CAM) has been widely popular as it highlights the specific sections of the MRI that influenced the model's predictions.[7] These insights are critical for allowing physicians to visually review and validate the model's predictions for better patient treatment outcomes.

Current diagnostic methods for GBM using AI have made limited advancements due to a lack of research funding until recent developments.[2] Existing tumor segmentation models are consistently high-performing but have a critical limitation in how they provide interpretable insights for medical professionals to support their clinical decisions. Certain key examples include models integrating Resnet-50 or Inception-ResNetV2 with Grad-CAM.[8,9] Notably, it is rather computationally heavy, requiring higher processing power, and will be difficult to optimize for clinical use in hospitals. Even current models, such as Vision Transformers ensemble models, would increase the computational complexity and substantial resources de-

DOI: 10.36838/v7i6.64

spite offering highly advanced feature extraction capabilities, which are crucial for GBM segmentation.[9] One major trend amongst most of the existing Grad-CAM models is their high accuracy, consistently remaining over 95%.[10-13] This highlights how accuracy is not the sole determinant and improved clinical applicability is crucial for the successful implementation of such models. Another limitation is the reliance on single or dual-modality modalities. A lack of data modalities can directly lead to performance degradation because these modalities provide complementary information to enhance overall detection and classification.[14] This limits the model's ability to fully analyze the morphological and vascular variations of GBM. Therefore, despite significant advancements, there is still a need for a multimodal, lightweight, and interpretable XAI model that still maintains high accuracy.

Fresh approach:

Our model, Neuroflux, brings unique capabilities to existing models by integrating performance with clinical usability. A lightweight and computationally efficient 2-dimensional U-Net CNN structure is the core of Neuroflux's unique approach to XAI. Neuroflux's multimodal approach of incorporating multiple MRI sequences, such as post-contrast T1-weighted (T1ce) and Fluid Attenuated Inversion Recovery (FLAIR), offers a more comprehensive representation of GBM tumor regions. This added layer of depth will improve Neuroflux's clinical relevance and usability. Current models also largely utilize Sparse Categorical Cross-Entropy or Binary Cross-Entropy loss, which are excellent functions but are generally more suited for classification tasks rather than segmentation.[8,9,15] Unlike these models, Neuroflux includes a unique hybrid loss function consisting of a categorical cross-entropy and a custom Dice loss. Combining both precision and optimization, Neuroflux is positioned as a comprehensive, scalable, and clinically-based XAI model. The goal of our research is to combine high performance with applicability and interpretability and effectively address current limitations in GBM segmentation. We will be discussing the evaluation results of Neuroflux to achieve three objectives — accurately predicting tumors, creating heatmaps through Grad-CAM, and assessing its overall interpretability. Through this study, we aim to draw further conclusions from Neuroflux's neural patterns and discuss these results in the context of current literature.

Methods

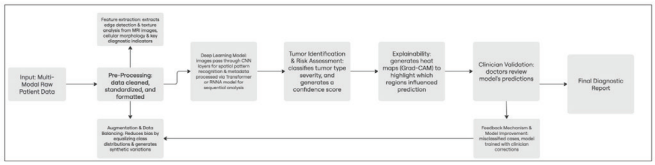

Figure 1: A visual workflow of Neuroflux's process in GBM segmentation. Preprocessing, feature extraction, and classifications of multi-modal patient data were performed using deep learning for tumor type and severity determination. Key features include Grad-CAM to support clinician verification and the feedback loop to fine-tune Neuroflux for generating more precise and reliable diagnostic reports.

Dataset and preprocessing:

We trained and assessed Neuroflux by leveraging the BraTS 2020 dataset, a well-known collection of brain MRI scans.[16,17,19,20,23] This dataset was adequate for our task because it includes multimodal MRI scans, such as native (T1-weighted), T1ce, T2-weighted, and T2 FLAIR scans. These various imaging modalities gave the model a solid, comprehensive training foundation that allowed it to capture a broad range of anatomical and pathological features. The information was gathered from 19 different institutions and clinics that utilized a range of imaging equipment and clinical procedures, guaranteeing a significant degree of variation within the dataset. Corresponding segmentation masks are used to identify tumor sub-regions, such as necrotic and non-enhancing tumor core, peritumoral edema, and enhancing tumor regions. This process is outlined in the visual workflow displaying Neuroflux's process in GBM segmentation (Figure 1). As part of the preprocessing processes, a MinMaxScaler was used to normalize intensity data, thus ensuring consistent feature representation for the model across various scans and modalities. A compromise had to be met to save computational expenses while maintaining crucial spatial features, so the photos were resized to 128x128. To conform to the model output format, the segmentation masks were transformed into categorical one-hot encoded representations. This step was necessary to ensure proper compatibility with the softmax activation function in the final layer. To enhance model generalization, rotation, flipping, and intensity shifting were applied to the training dataset.

U-Net model architecture and modifications:

The U-Net model used for tumor segmentation was adopted and adapted from a baseline model available on Kaggle titled Brain Tumor Segmentation Using U-Net, developed by Zeeshan Latif.[20] Neuroflux retains the classic encoder-decoder architecture with skip connections, a design choice to retain spatial information lost during down-sampling. However, key architectural modifications were made to the original model, some of which include incorporating dropout layers with a dropout rate of 0.2 following the bottleneck layer, addressing the issue of overfitting. To preserve consistency in the intensity distribution across many MRI images, we used MinMax scaling over standardization for normalization. This choice was made to handle variations in contrast and brightness across different scanners. A hybrid loss function allows us to combine categorical cross-entropy with a custom Dice loss to optimize segmentation for imbalanced classes. While Dice loss improved segmentation overlap for small tumor patches, categorical cross-entropy assisted Neuroflux in learning pixel-wise classifications. Finally, a softmax activation function was used in the final layer to predict pixel-wise class probabilities for the four segmentation classes, ensuring mutually exclusive predictions at the pixel level. Convolutional layers were initialized using the He normal initialization to enhance convergence speed and weight stability.

Training and evaluation metrics:

The dataset was divided into three subsets using patient-wise splitting to prevent data leakage, thus ensuring that all MRI scans from a single patient remain within a single subset, to avoid any potential overlap between training, validation, and test sets. The data was proportioned as follows: 68% of the data for training, 20% of the data for validation, and 12% of the data for testing, to ensure a diverse representation across different institutions. The Adam optimizer, which has a learning rate of 0.001, was one of our training hyperparameters. It was selected for its effective convergence and adaptive learning rates, which aided Neuroflux in preventing saddle points. To try and balance training stability and memory efficiency, a batch size of four was chosen. Neuroflux was trained for 25 epochs with early stopping to prevent overfitting, ensuring that training terminated when validation loss no longer improved. When validation loss plateaued, a ReduceLROnPlateau callback was used to lower the learning rate by a factor of 0.2 to optimize performance at later training phases. To assess segmentation quality, we computed the following 5 metrics in our evaluation. Our first metric was accuracy, which measures the overall proportion of correctly classified pixels. The second metric used was the Dice coefficient, which evaluates the overlap between predicted and ground truth segmentations, making it particularly useful for measuring performance on small tumor regions. Our third metric is precision and sensitivity, where precision, specifically positive predictive value, assesses how many predicted tumor pixels were tumors, while sensitivity, or recall, measures how many actual tumor pixels were correctly identified. Our fourth metric is specificity, which gauges the true negative rate and makes sure Neuroflux doesn't mistakenly identify tumors in healthy tissue. Finally, Intersection over Union (IoU) was our fifth metric used in Neuroflux. The IoU allowed us to evaluate the segmentation overlap consistency by comparing intersection regions to the total region union. To provide a detailed analysis of Neuroflux's performance on various tumor subtypes, custom Dice coefficients were calculated independently for necrotic, edema, and enhancing tumor regions.

Model interpretability with Grad-CAM:

Our most significant modification to the baseline model was enhancing interpretability by implementing Grad-CAM, which generates heat maps that highlight regions of the image that contribute the most to Neuroflux's parameter tuning and, consequently, to its predictions. Through more transparent tumor region localization, this visualization technique enables clinicians to validate model decisions. The target layer for Grad-CAM was set at the deepest convolutional layer, i.e. conv5, ensuring that high-level features influenced the generated heatmaps. Our Grad-CAM implementation included target class selection, where Grad-CAM was applied separately to each tumor class, visualizing attention for necrotic core, peritumoral edema, and enhancing tumor regions. Heatmap overlays were used to superimpose the generated heatmaps on original MRI slices to verify whether Neurofluxl focused on relevant tumor structures. The heatmaps were subjected to thresholding, which eliminated low-importance activations and only highlighted the most important regions.

Result and Discussion

This section will review Neuroflux's results based on an array of training and evaluation metrics. The discussion will be based on our three key objectives mentioned previously, with our analysis of Neuroflux's significance in the current neuro-oncology landscape.

Features	Neuroflux	Existing GBM Segmentation Models			
		ResNet-50 + Grad-CAM[8]	Inception-ResNetV2 + Grad-CAM[9]	IVX16[21]	Vision Transformers Ensemble[22]
Architecture	Modified U-Net	ResNet50 CNN	Inception-ResNetV2 CNN	IVX16 Architecture	Vision transformers
Loss Function	Categorical cross-entropy and Dice loss	Binary cross-entropy loss	Categorical cross-entropy loss	Binary cross-entropy loss	Dice coefficient loss
Accuracy	99.01%	96.23%	98.52%	96.94%	86.6%
Computational Efficiency	Lightweight	Heavy	Heavy	Heavy	Very heavy
Explainability	Grad-CAM	Grad-CAM	Grad-CAM	None	None
Target Use	Tumor diagnosis and clinical interpretability	Tumor diagnosis and clinical interpretability	Tumor diagnosis and clinical interpretability	Tumor diagnosis	Tumor diagnosis

Figure 2: A table consisting of Neuroflux's key features compared to other GBM segmentation models. The consolidated table compares the configurations of various segmentation models and the accuracy achieved by each model. Additionally, it breaks down each model's intended use as not all models include Grad-CAM integration to provide clinical interpretability.

The performance and configuration of Neuroflux are compared to four of the most commonly used and referenced GBM segmentation models (Figure 2). Neuroflux achieves the highest pixelwise accuracy (meaning it correctly classifies the greatest proportion of pixels as non-tumor, necrotic, edema, or enhancing pixels, which is critical for the implementation of Grad-CAM) out of the five total models while remaining the computationally least expensive. This is due to the modified U-Net that Neuroflux is based on and its unique use of a hybrid loss function utilizing Categorical Cross-Entropy and Dice loss.

Objective 1: Accurately predicting tumors:

The modified U-Net model for brain tumor segmentation, enhanced with Grad-CAM for interpretability, demonstrated strong performance across all key evaluation metrics. By incorporating Grad-CAM into the model, Neuroflux not only identifies tumors but also produces interpretable results. Additionally, Neuroflux leverages Dice Loss, addressing class imbalances by focusing on overlapping regions between predicted and actual masks. Below, we summarize the results and their implications for Neuroflux's reliability and explainability.

Neuroflux achieved an accuracy of 99.01%, showing that almost all pixel predictions, both tumorous and non-tumorous, were correct. With a loss value of 0.0319, it exhibited minimal deviation between predicted and true labels, indicating effective learning of the segmentation task. Neuroflux also recorded a precision of 99.11%, highlighting its ability to minimize false positives in its predictions. At 98.85%, the sensitivity shows

DOI: 10.36838/v7i6.64

Neuroflux's ability to correctly identify actual tumor regions. A specificity of 99.7% underscores Neuroflux's capability to accurately recognize non-tumor regions.

Neuroflux's segmentation accuracy surpasses many existing deep learning-based tumor segmentation models. For example, a nnU-Net model achieved an accuracy range of 94% to 97%.[18] Similarly, DeepMedic, an established model for brain tumor segmentation, reports a mean dice score of 0.82, performing lower than Neuroflux.[19] This demonstrates the potential for the integration of Grad-CAM while maintaining high segmentation accuracy through a U-Net framework.

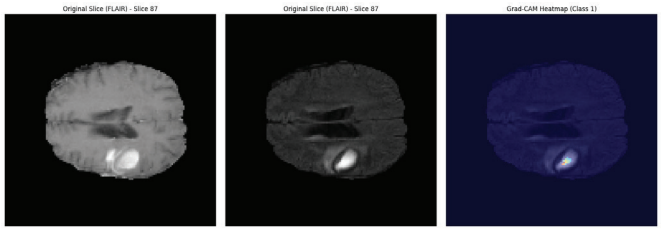

Figure 3: BraTS2020 Dataset Slice 95 Shown in T1ce and FLAIR modalities. T1ce (left) shows the specific tumor boundaries, while FLAIR (right) shows the full brain tissue damage. These pixels are what gets fed into the U-Net model for it to segment images and identify distinctions or abnormalities.

Another key advantage of Neuroflux is its incorporation of multimodal MRI sequences, utilizing both T1ce and FLAIR sequences. This captures both structural and vascular characteristics of GBM, creating a more comprehensive representation of tumor morphology. T1ce uses high contrast to distinguish abnormalities, while FLAIR sequences suppress fluid signals and make it easier to identify lesions, particularly lesions in the brain's white matter. The use of both modalities allows the model to differentiate between tumor subregions more effectively, reducing misclassification and improving model accuracy. Additionally, the Grad-CAM visualizations obtained from multiple modalities ensure that the highlighted regions are relevant between both modalities. An example of each modality from the BraTS2020 dataset underscores each version's unique benefits (Figure 3). In the T1ce image, the tumor-enhanced area appears in higher contrast, pointing out the tumor's boundaries. In the FLAIR image, the areas where brain tissue has swelled around the tumor are highlighted, revealing the extent of the tumor's damage.

Objective 2: Creating heatmaps through Grad-CAM:

Building upon the strong segmentation results, Grad-CAM was implemented to provide interpretability and insights into Neuroflux's decision-making process. This integration visualizes the regions of the input image that strongly influence its predictions, enabling clinicians to be able to validate Neuroflux's outputs. By highlighting areas of interest in tumor segmentation, Grad-CAM serves as a critical tool for understanding and verifying Neuroflux's behavior. Unlike previous applications of Grad-CAM that primarily focus on classification, our approach applied Grad-CAM at multiple levels of the U-Net architecture. Through this, the visualization of feature activations across different spatial resolutions is carried out and further improves the accuracy of highlighted regions. The use of multimodal MRI sequences allows Neuroflux to distinguish between different tissue types and reduce its dependency on a single modality, a key limitation in many existing models.

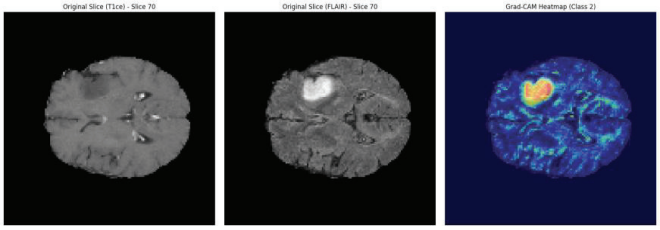

Figure 4: BraTS2020 Dataset Case 200 Slice 70 Enhanced with Grad-CAM. A region's color warmth corresponds to that region's effect on the algorithm's output. Tumor regions are shown in high contrast with warm colors, and you can see where the algorithm detected a tumor in the upper left of the brain scan. The clarity with which Grad-CAM displays the tumor region highlights the implementation's success.

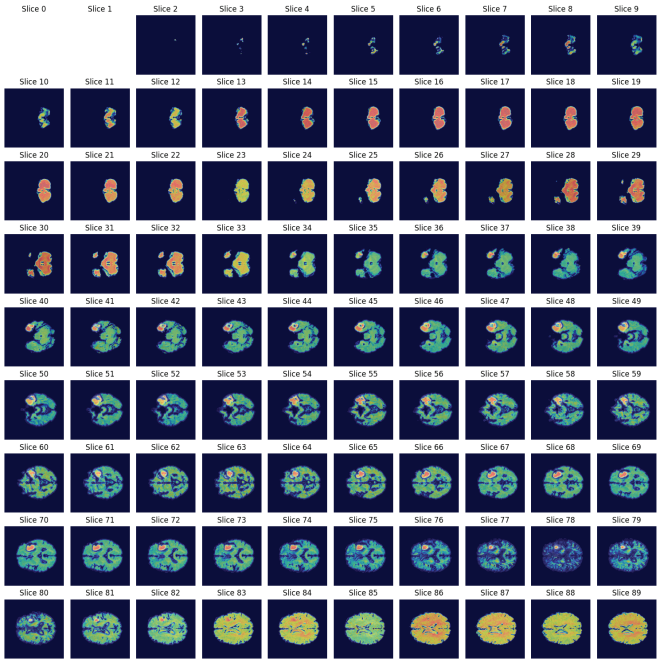

Figure 5: BraTS2020 Dataset Case 200 All Slices Enhanced with Grad-CAM. Slices 60–80 show the best results for interpretability, highlighting the tumor region's location within the entire brain. The grid shows Grad-CAM's learning process and how it learns as it gets more angles from the brain. Professionals may choose to look at diagnoses like this to choose the most representative angle(s) themselves.

A piece of data (Figure 4) is extracted from the dataset, and the portion of the brain that had the most significant impact on its diagnosis is highlighted in 3 different color maps. When another piece of data is extracted and analyzed from every MRI angle (Figure 5), each angle demonstrates the portion of the brain that had the most significant impact on its diagnosis in a heatmap representation. The red and yellow areas indicate strong activations, signifying how Neuroflux focused on these regions when classifying the presence of a tumor, while green and blue areas show moderate influence and dark blue or black regions had little to no impact. In the earlier slices from Slice 0 to Slice 10, minimal activation is present, whereas Slices 65 to Slice 75 show strong activations in localized areas, indicating

likely tumor presence. By interpreting the heatmaps, this visualization by Neuroflux will help pathologists assess whether the model is correctly identifying tumor-related features.

Objective 3: Assessing the interpretability:

The development of Neuroflux emphasizes the need to balance AI performance and explainability. While high segmentation accuracy is essential, combating the ethical implications that come with AI decision-making would improve trust in AI-driven decisions. The integration of Grad-CAM not only adds to Neuroflux's explainability but also ensures that ethical XAI practices are followed, and results are verifiable and transparent. This thereby allows medical practitioners to examine Neuroflux's decision-making. This is essential as AI algorithms are still in the early stages of development and currently cannot be fully trusted. With imperative decisions such as GBM diagnosis, it is essential that clinicians can confirm the conclusion of an algorithm.[23]

While Grad-CAM provides valuable transparency into an algorithm's decision-making process, its current implementation is limited in terms of precision. On certain occasions, the generated heat map may highlight a broad area rather than the exact boundaries of a tumor. Future improvements such as implementing a Layer-wise Relevance Propagation (LRP) or SHapley Additive exPlanations (SHAP). An LRP will redistribute Neuroflux's output backward through a neural network's layers to determine which features contributed most to the prediction. SHAP, which uses game theory principles to assign each feature a contribution score, could also offer more granularity via pixel-by-pixel importance scores.[24] Considering the weight of a GBM diagnosis, providing neurologists and oncologists with that additional level of precision is vital.

The main strength of Neuroflux is its lightweight design, ensuring high performance without excessive computational demand. To balance accuracy, efficiency, and model stability, carefully optimized dropout rates, a MinMax scaler to scale features, and a hybrid loss strategy were utilized. A contrast witnessed while creating Neuroflux was between model performance and computational efficiency. One of the biggest challenges in deploying AI models for oncological use lies in a model's computational efficiency. Most of the current high-performing models are very intensive in terms of resources, therefore making them impractical to use in a hospital setting. Neuroflux resolves this constraint through the use of a U-Net architecture.

Conclusion

Overall, Neuroflux demonstrates how AI can bridge gaps in GBM diagnosis through diverse mechanisms such as combining high-accuracy segmentation and the application of Grad-CAM with a U-Net framework to create transparency, confidence, and integration in clinical environments.[25,26] This combination ensures segmentation accuracy while surmounting a major flaw of "black box" AI models of explainability. By integrating Grad-CAM into a U-Net framework, Neuroflux yields higher clinical interpretability, enabling clinicians to develop a deeper understanding of the model's predictions.

This study investigates the possibilities of how AI-powered solutions can revolutionize the diagnostics of diseases such as GBM. Neuroflux reduces human effort, and interobserver variation, and streamlines the overall diagnostics process. Considering GBM's fast-developing nature, early and accurate interventions are crucial in improving patient outcomes.[27] In contrast to the observer-dependent manual segmentation process, Neuroflux is capable of providing reliable and automated outputs while maintaining interpretability through Grad-CAM visualization. This visual interpretability with high performance closes the gap between AI-driven automation and clinical decision-making, ensuring that segmentation outcomes remain aligned with anatomical structures and radiological evaluations.

Neuroflux demonstrates how performance-driven AI does not have to be clinically irrelevant or inaccurate, breaking the trade-off between performance and usability. Through specialized model adjustments, Neuroflux's overall performance was improved from the baseline model without compromising on robustness. In particular, the optimized dropout rate prevented overfitting while the MinMax scaling normalized intensity variations across scans. Moreover, the hybrid loss function maintains pixel-level accuracy along with overall structural coherence, enhancing Neuroflux's flexibility across varying imaging environments.

Future research could focus on expanding the model beyond GBM. This could be implemented by further integration of other image modalities such as Computed Tomography (CT) and Positron Emission Tomography (PET). Training Neuroflux on CT and PET scans will further increase its generalizability across a variety of diagnostic procedures. Additionally, optimizing computational efficiency for real-time clinical integration will allow for a more accurate diagnostic tool. Moreover, greater validation of Neuroflux with larger, more heterogeneous datasets will continue to challenge its robustness and generalizability.[28] This will facilitate wider oncological use cases, making AI-assisted diagnostics more versatile and accessible beyond GBM.

Acknowledgments

We would like to extend our heartfelt gratitude to The Junior Academy at the New York Academy of Sciences for providing us with the platform and opportunity to collaborate. Moreover, we would also like to acknowledge the BraTS 2020 dataset and the "Brain Tumor Segmentation using U-Net" model from Zeeshan Latif, which were invaluable in forming the foundation of Neuroflux's development.

Availability of data and materials:

The full code used for Neuroflux is available at https://github.com/neuroflux-ai/neuroflux. This includes all our preprocessing, training, and evaluation scripts with further documentation for reproducibility. The repository is shared under the Apache 2.0 license agreement. More information about Neuroflux and its impacts are also available at https://www.neurofluxai.org/.

DOI: 10.36838/v7i6.64

References

1. Urbańska, K.; Sokołowska, J.; Szmidt, M.; Sysa, P. Review Glioblastoma multiforme – an overview. *Współczesna Onkologia* 2014, *5*, 307–312. https://doi.org/10.5114/wo.2014.40559. 2. Grochans, S.; Cybulska, A.M.; Simińska, D.; Korbecki, J.; Kojder, K.; Chlubek, D.; Baranowska-Bosiacka, I. Epidemiology of Glioblastoma Multiforme–Literature Review. *Cancers* 2022, *14*, 2412. https://doi.org/10.3390/cancers14102412
3. Bhandari, A.; Koppen, J.; Agzarian, M. Convolutional neural networks for brain tumour segmentation. *Insights Into Imaging* 2020, *11*(1). https://doi.org/10.1186/s13244-020-00869-4.
4. Perkuhn, M.; Stavrinou, P.; Thiele, F.; Shakirin, G.; Mohan, M.; Garmpis, D.; Kabbasch, C.; Borggrefe, J. Clinical evaluation of a multiparametric deep learning model for glioblastoma segmentation using heterogeneous magnetic resonance imaging data from clinical routine. *Investigative Radiology* 2018, *53* (11), 647–654. https://doi.org/10.1097/rli.0000000000000484.
5. Chang, J.; Zhang, L.; Gu, N.; Zhang, X.; Ye, M.; Yin, R.; Meng, Q. A mix-pooling CNN architecture with FCRF for brain tumor segmentation. *Journal of Visual Communication and Image Representation* 2018, *58*, 316–322. https://doi.org/10.1016/j.jvcir.2018.11.047.
6. Chang, J.; Zhang, L.; Gu, N.; Zhang, X.; Ye, M.; Yin, R.; Meng, Q. A mix-pooling CNN architecture with FCRF for brain tumor segmentation. *Journal of Visual Communication and Image Representation* 2018, *58*, 316–322. https://doi.org/10.1016/j.jvcir.2018.11.047.
7. Kumar, K.; Jyoti, K. Recent Advancements in Grad-CAM and variants: Enhancing Brain Tumor Detection, Segmentation, and Classification. *Research Square (Research Square)* 2024. https://doi.org/10.21203/rs.3.rs-5485128/v1.
8. M, M. M.; R, M. T.; Kumar, V., V.; Guluwadi, S. Enhancing brain tumor detection in MRI images through explainable AI using Grad-CAM with Resnet 50. *BMC Medical Imaging* 2024, *24* (1). https://doi.org/10.1186/s12880-024-01292-7.
9. Talaat, F. M.; Gamel, S. A.; El-Balka, R. M.; Shehata, M.; ZainEldin, H. Grad-CAM Enabled Breast Cancer Classification with a 3D Inception-ResNet V2: Empowering Radiologists with Explainable Insights. *Cancers* 2024, *16* (21), 3668. https://doi.org/10.3390/cancers16213668.
10. Hossain, S.; Chakrabarty, A.; Gadekallu, T. R.; Alazab, M.; Piran, Md. J. Vision transformers, ensemble model, and transfer learning leveraging explainable AI for brain tumor detection and classification. *IEEE Journal of Biomedical and Health Informatics* 2023, *28*(3), 1261–1272. https://doi.org/10.1109/jbhi.2023.3266614.
11. Kumar, K. K.; Dinesh, P. M.; Rayavel, P.; Vijayaraja, L.; Dhanasekar, R.; Kesavan, R.; Raju, K.; Khan, A. A.; Wechtaisong, C.; Haq, M. A.; Alzamil, Z. S.; Alhussen, A. Brain tumor identification using data augmentation and transfer learning approach. *Computer Systems Science and Engineering* 2023, *46* (2), 1845–1861. https://doi.org/10.32604/csse.2023.033927.
12. Anaya-Isaza, A.; Mera-Jiménez, L.; Verdugo-Alejo, L.; Sarasti, L. Optimizing MRI-based brain tumor classification and detection using AI: A comparative analysis of neural networks, transfer learning, data augmentation, and the cross-transformer network. *European Journal of Radiology Open* 2023, *10*, 100484. https://doi.org/10.1016/j.ejro.2023.100484.
13. Rahman, T.; Islam, M. S. MRI brain tumor detection and classification using parallel deep convolutional neural networks. *Measurement Sensors* 2023, *26*, 100694. https://doi.org/10.1016/j.measen.2023.100694.
14. Khan, M. A.; Khan, A.; Alhaisoni, M.; Alqahtani, A.; Alsubai, S.; Alharbi, M.; Malik, N. A.; Damaševičius, R. Multimodal brain tumor detection and classification using deep saliency map and improved dragonfly optimization algorithm. *International Journal of Imaging Systems and Technology* 2022, *33* (2), 572–587. https://doi.org/10.1002/ima.22831.
15. S, Y. K.; Jeya, J. J.; R, M. T.; Khan, S. B.; Alzahrani, S.; Alojail, M. Explainable lung cancer classification with ensemble transfer learning of VGG16, Resnet50 and InceptionV3 using grad-cam. *BMC Medical Imaging* 2024, *24* (1). https://doi.org/10.1186/s12880-024-01345-x.
16. Menze, B. H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R.; Lanczi, L.; Gerstner, E.; Weber, M.-A.; Arbel, T.; Avants, B. B.; Ayache, N.; Buendia, P.; Collins, D. L.; Cordier, N.; Corso, J. J.; Criminisi, A.; Das, T.; Delingette, H.; Demiralp, C.; Durst, C. R.; Dojat, M.; Doyle, S.; Festa, J.; Forbes, F.; Geremia, E.; Glocker, B.; Golland, P.; Guo, X.; Hamamci, A.; Iftekharuddin, K. M.; Jena, R.; John, N. M.; Konukoglu, E.; Lashkari, D.; Mariz, J. A.; Meier, R.; Pereira, S.; Precup, D.; Price, S. J.; Raviv, T. R.; Reza, S. M. S.; Ryan, M.; Sarikaya, D.; Schwartz, L.; Shin, H.-C.; Shotton, J.; Silva, C. A.; Sousa, N.; Subbanna, N. K.; Szekely, G.; Taylor, T. J.; Thomas, O. M.; Tustison, N. J.; Unal, G.; Vasseur, F.; Wintermark, M.; Ye, D. H.; Zhao, L.; Zhao, B.; Zikic, D.; Prastawa, M.; Reyes, M.; Van Leemput, K. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging* 2014, *34* (10), 1993–2024. https://doi.org/10.1109/tmi.2014.2377694.
17. Bakas, S.; Akbari, H.; Sotiras, A.; Bilello, M.; Rozycki, M.; Kirby, J. S.; Freymann, J. B.; Farahani, K.; Davatzikos, C. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data* 2017, *4* (1). https://doi.org/10.1038/sdata.2017.117.
18. Bakas, S.; Akbari, H.; Sotiras, A.; Bilello, M.; Rozycki, M.; Kirby, J.; Freymann, J.; Farahani, K.; Davatzikos, C.Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM Collection. Cancer Imaging Arch. 2017. https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q.
19. Bakas, S.; Akbari, H.; Sotiras, A.; Bilello, M.; Rozycki, M.; Kirby, J.; Freymann, J.; Farahani, K.; Davatzikos, C.Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG Collection. Cancer Imaging Arch. 2017.https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF.
20. Latif, Z. *Brain Tumor Segmentation using U-Net*. Kaggle. https://www.kaggle.com/code/zeeshanlatif/brain-tumor-segmentation-using-u-net/notebook (accessed 2025-02-17).
21. Liu, H.; Dowdell, B.; Engelder, T.; Pulmano, Z.; Osa, N.; Barman, A. *Glioblastoma Tumor Segmentation using an Ensemble of Vision Transformers*. arXiv. https://arxiv.org/abs/2312.11467
22. Hossain, S.; Chakrabarty, A.; Gadekallu, T.; Alazab, M.; Piran, J. Vision Transformers, Ensemble Model, and Transfer Learning Leveraging Explainable AI for Brain Tumor Detection and Classification. *IEEE Journal of Biomedical and Health Informatics* 2024, *28* (3), 1261–1272. https://doi.org/10.1109/JBHI.2023.3266614.
23. Bakas, S.; Reyes, M.; Battistella, E.; Chandra, S.; Estienne, T.; Fidon, L.; Vakalopoulou, M.; Sun, R.; Al, E.; Deutsch, É.; Robert, C.; Paragios, N.; Menze, B. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. *arXiv (Cornell University)* 2018. https://doi.org/10.48550/arxiv.1811.02629.
24. Chereda, H.; Leha, A.; Beißbarth, T. Stable feature selection utilizing Graph Convolutional Neural Network and Layer-wise Relevance Propagation for biomarker discovery in breast cancer. *Artificial Intelligence in Medicine* 2024, *151*, 102840. https://doi.org/10.1016/j.artmed.2024.102840.

25. Yousef, R., Khan, S., Gupta, G., Siddiqui, T., Albahlal, M., Alajlan, A., & Haq, A. (2023, May 4). *U-Net-Based Models Optimal MR Brain Image Segmentation*. National Library of Medicine. Retrieved February 9, 2025, from https://pmc.ncbi.nlm.nih.gov/articles/PMC10178263/
26. Zheng, P., Zhu, X., & Guo, W. (2022). Brain tumor segmentation based on an improved U-Net. *BMC Medical Imaging*, *22*(1). https://bmcmedimaging.biomedcentral.com/articles/10.1186/s12880-022-00931-1
27. Vossough, A.; Khalili, N.; Familiar, A.; Gandhi, D.; Viswanathan, K.; Tu, W.; Haldar, D.; Bagheri, S.; Anderson, H.; Halder, S.; Storm, P.; Resnick, A.; Ware, J.; Nabavizadeh, A.; Kazerooni, A. *Training and comparison of NNU-Net and DeepMedic methods for autosegmentation of pediatric brain tumors*. National Library of Medicine. https://pubmed.ncbi.nlm.nih.gov/38724204/
28. Zhang, H., & Ogasawara, K. (2023, March 10). *Grad-CAM-Based explainable artificial intelligence related to medical text processing*. National Library of Medicine. https://pubmed.ncbi.nlm.nih.gov/37760173/

Authors

Eve Ang is a Year 6 student at Saint Joseph's Institution in Singapore. She has earned global recognition and international funding for her health innovations and is interested in biomedical sciences and healthcare sustainability. She is particularly passionate about oncological informatics and plans to pursue a career in biomedical entrepreneurship.

Darius Oltei is a junior at "Octavian Goga" National College in Sibiu, Romania. He has a strong interest in computer science, machine learning, and data science, with a focus on applying these fields to real-world challenges. He is passionate about solving complex problems and applying his knowledge in practical ways.

Pranav Vedula is a junior at Colonial Forge High School in the United States. He is passionate about computer science and business, particularly how entrepreneurship enables AI solutions to be accessible and implemented. He has conducted relevant research and is planning to work in computer science and business administration.

Shriya Viswanathan is a junior at Plano Senior High School in the United States. She is passionate about AI, computational biology, and the intersection of technology and entrepreneurship. Having conducted research in computational sciences and worked with AI-driven innovations, she is dedicated to using technology to drive real-world impact.

DOI: 10.36838/v7i6.64