DGG-XNet: A Hybrid Deep Learning Framework for Multi-Class Brain Disease Classification with Explainable AI

Sumshun Nahar Eity[1][0009−0003−4614−3883], Mahin Montasir Afif[1][0009−0003−8499−2624], Tanisha Fairooz[1][0009−0004−0898−1381], Md. Mortuza Ahmmed[1][0000−0002−4735−571X], and Md Saef Ullah Miah[1][0000−0003−4587−4636]

American International University-Bangladesh, Dhaka-1229, Bangladesh
22-48227-2@student.aiub.edu, 22-46573-1@student.aiub.edu,
22-49347-3@student.aiub.edu, mortuza@aiub.edu, saef@aiub.edu

Abstract. Accurate diagnosis of brain disorders such as Alzheimer's disease and brain tumors remains a critical challenge in medical imaging. Conventional methods based on manual MRI analysis are often inefficient and error-prone. To address this, we propose DGG-XNet, a hybrid deep learning model integrating VGG16 and DenseNet121 to enhance feature extraction and classification. DenseNet121 promotes feature reuse and efficient gradient flow through dense connectivity, while VGG16 contributes strong hierarchical spatial representations. Their fusion enables robust multiclass classification of neurological conditions. Grad-CAM is applied to visualize salient regions, enhancing model transparency. Trained on a combined dataset from BraTS 2021 and Kaggle, DGG-XNet achieved a test accuracy of 91.33%, with precision, recall, and F1-score all exceeding 91%. These results highlight DGG-XNet's potential as an effective and interpretable tool for computer-aided diagnosis (CAD) of neurodegenerative and oncological brain disorders.

Keywords: Alzheimer's disease · Brain Tumors · DenseNet121 · VGG16 · DGG-XNet · Grad-CAM·

1 Introduction

Brain disorders such as Alzheimer's disease together with brain tumors are substantial global health issues. Millions of patients are impacted, and medical practitioners face challenges in diagnosing them [2]. Studies have demonstrated that, depending on the patient's age and the brain region involved, 52% of patients had Alzheimer's Disease Neuropathologic Change (ADNC) in the glioblastoma-adjacent cortex tissue [9]. According to research, there is presently a demand for enhanced computational diagnostics to increase diagnostic speed due to the rising number of people with Alzheimer's disease (246 million globally, 39 million of whom are critical) and brain tumors. According to the World Health Organization [24], since over 3 billion individuals worldwide suffer from various brain diseases, early detection methods become crucial.

In order to detect diseases medical personnel perform interpretive analysis of Magnetic Resonance Imaging (MRI) which scans the human body, even though it's a revolutionary invention it is still a slow process and contains manual errors. Although in the current circumstances, MRI is the only leading medical tool for discovering brain diseases that involve neurodegeneration and cancer disorders. However, acquiring the MRI report requires people to see the doctor again to obtain the findings, which is both costly and time-consuming. For better medical outcomes, timely, accurate diagnosis requires both accuracy and efficiency, as evidenced by the number of diagnosed patients.

The goal of this research is the development of an efficient and quick detection model using MRI technology in recognition of Alzheimer's disease and brain tumors at an early stage. In combining advanced imaging methods with machine learning, this study hopes to improve diagnostic accuracy and thus patient outcomes and treatment plans.

CNNs have shown remarkable effectiveness in picture classification tasks, single designs usually have trouble applying effectively across a range of neurological disorders. To face this circumstance, this investigation presents DGG-XNet, it is a novel hybrid deep learning model based on VGG16 and DenseNet121, which fully utilize their characteristic extractions. Regarding DenseNet121 [10], its feature reuse is greatly improved by dense connections generated via the usage of dense layers, while VGG16's [23] deep hierarchical design keeps robust spatial characteristics useful for medical image-related jobs fiercely. DGG-XNet is designed to improve the multi-class classification efficiency of brain tumors and Alzheimer's disease by combining these two models.

2   Literature Review

2.1   Alzheimer's disease

Pursuit of Alzheimer's disease (AD) diagnosis becomes an increasing fear for persons as they age. The cellular breakdown from Alzheimer's disease leads people to lose memory of familiar people along with the ability to learn new data or recognize their relatives [12]. The disease affects the ability to determine which direction people should go and their ability to recognize the environment. Eating is combined with breathing and coughing as abilities that fade away during the advanced stages of disease progression. AD and VD, share sufficient overlapping symptoms that make diagnosing AD particularly challenging both in behavioral and psychiatric domains as well as in language and memory dysfunction [11,5]. Early and accurate identification of AD and monitoring of its progression will greatly benefit patient care as well as treatment and prevention.

2.2   Brain Tumor

A brain tumor stands as an extremely dangerous medical condition that threatens the safety of the brain. Brain tumors develop regularly because the existing

vein and nerve damage in the brain leads to this condition. The progression of tumor growth determines whether the patient loses partial or complete sight to this condition [25]. The development of brain tumors also depends on serious myopia cases as well as ethnic characteristics and genetic background [8]. The condition is caused by insufficient blood circulation and prevents new nerve blood vessels from growing. Rapid and automated, the discovery of early diagnostic techniques has become crucial in modern developed societies because of their advanced requirements.

2.3   Deep Learning and Brain Diseases

Deep learning techniques have become widely used in medical image classification by providing effective approaches for classifying diseases in brain MRI scans. A study by [7] shows an automated system for the early detection of Alzheimer's disease using transfer learning on multi-class classification of brain MRI images. The proposed model achieved 91.70% accuracy. Similar study by [22] using two MRI datasets-containing 6400 and 6330 images-employed a neural network classifier with VGG16 as the feature extractor. The model achieved accuracy, precision, recall, AUC, and F1-score of 90.4%, 0.905, 0.904, 0.969, and 0.904, respectively, on the first dataset, and 71.1%, 0.71, 0.711, 0.85, and 0.71 on the second dataset. This study by [18] shows an approach to optimize CNNs using the Fuzzy Gravitational Search Algorithm, focusing on key parameters like image block size, filter number, and filter size. Applied to the ORL and Cropped Yale databases, the optimized CNN outperformed non-optimized models. In AD classification, VoxCNN achieved 79% accuracy and 88% AUC, while ResNet achieved 80% accuracy and 87% AUC.

A research by [20] proposes a WHHO-based DeepCNN model for brain tumor detection using MRI images. It combines Whale Optimization and Harris Hawks Optimization with DeepCNN for segmentation and feature extraction. The model achieved 0.816 accuracy, 0.791 specificity, and 0.974 sensitivity, outperforming other methods. Another study by [14] investigates a deep learning-based approach for tumor classification, combining whole slide images and MRI scans. The model was evaluated on the 2020 Computational Precision Medicine Challenge in a 3-class unbalanced classification task. It achieved impressive performance with a balanced accuracy of 0.913. Another research paper by [17] proposes an automatic brain tumor detection method using genetic algorithm for image s and support vector machine (SVM) for classification of brain MRI images. The model uses CNN with Discrete Wavelet Transform (DWT) and classifies tumors as Normal vs. Not Normal, achieving 85% accuracy.

Although a good number of work has been done in multi-class brain disease classification, challenges such as accuracy, dataset imbalance, and identifying the most reliable regions of images for classification tasks still remain. To address these issues, this study proposes the DGG-XNet model, which combines VGG16 and DenseNet121 to enhance accuracy. Additionally, the model incorporates two Explainable AI (XAI) techniques including Grad-CAM and Integrated Gradients, to make its predictions more transparent and interpretable,

thereby improving its practical applicability in real-world medical imaging. Table 1 presents a comparison of accuracy among discussed existing works.

Table 1: Accuracy Comparison of Previous Works

Reference	Model	Dataset	Accuracy
[7]	Transfer learning	MRI images	91.70%
[22]	VGG16+feature extractor	2 MRI datasets	90.4%, 71.1%
[18]	VoxCNN, ResNet	ORL, Cropped Yale	79%, 80%
[20]	WHHO+DeepCNN	MRI images	81.6%
[14]	Ensemble model	TCIA+CBICA MRI images	91.3%
[17]	CNN+DWT, SVM	Brain Tumor Dataset	85%

3 Methodology

In this section, we describe our proposed methodology. An extensive review of earlier studies in the area of neuroimaging-based disease classification was conducted before this study. Based on the insights, a dual-path hybrid model was constructed and trained on a balanced dataset. The methodology pipeline is summarized in Fig. 1.
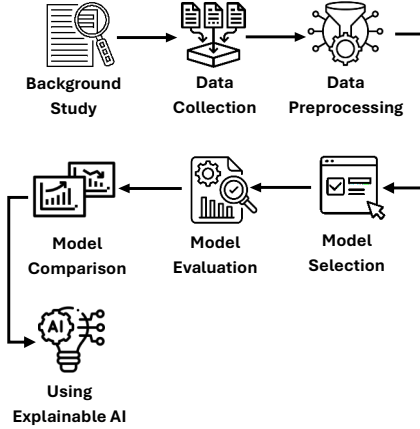

Fig. 1: Methodology Overview

3.1 Dataset Description

The study utilized two publicly available datasets. By combining "BraTS 2021 Task 1 dataset" [3,16,4] which contains MRI volumes for brain tumor classifi-

cation and "alzheimers-dataset-4-class-of-images" which contains 2D axial brain images across four categories: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented from Kaggle were collected and merged.

3.2   Data Preprocessing

Images were resized to 224×224×3 and normalized for model compatibility. From BraTS 2021 Task 1, Only the T1-weighted modality was used in this study. 2D slices were extracted from the 3D volumes by selecting evenly spaced axial slices to preserve critical spatial information. For the Alzheimer's dataset, images were pre-processed and resized uniformly. To prevent model bias due to class imbalance, the dataset was equalized using down sampling. The number of samples in each class was adjusted to match the minority class, resulting in 500 samples per class.

3.3   Dataset Split

The dataset was partitioned into training (70%), validation (20%), and testing (10%) sets using stratified sampling. This ensured that each subset retained balanced class distribution. Table 2 presents the distribution of the dataset in different sets.

Table 2: Dataset Distribution

Class	Train	Validation	Test
Tumour	350	100	50
Normal	350	100	50
Alzheimer's	350	100	50
Total	1050	300	150

3.4   Proposed Model

This study proposes a hybrid deep learning architecture that combines the feature extraction power of two widely used convolutional neural networks: VGG16 and DenseNet121. Both models were initialized with pre-trained ImageNet [6] weights to benefit from transfer learning, and their convolutional layers were initially frozen to retain learned representations.

Each input image is passed through both VGG16 and DenseNet121 backbones. Feature maps are extracted using Global Average Pooling [15] from each model, and then concatenated to form a unified feature vector:

$$F = \text{Concat}\left(\text{GAP}(\text{VGG16}(x)),\ \text{GAP}(\text{DenseNet121}(x))\right) \qquad (1)$$

where x is the input image and F is the fused feature vector.

The merged features are passed through a series of fully connected layers with ReLU activation [1] functions and dropout for regularization. The final output layer uses the softmax activation [19] to classify the input into one of three categories: Tumour, Normal, or Alzheimer's.

$$\hat{y} = \text{softmax}(W_2 \cdot \text{ReLU}(W_1 \cdot F + b_1) + b_2) \tag{2}$$

The model is compiled using the Adam optimizer [13] with a learning rate of 0.0001. Adam is an adaptive optimizer that updates parameters using first and second moment estimates of the gradients:

$$\theta_t = \theta_{t-1} - \frac{\alpha}{\sqrt{\hat{v}_t} + \epsilon} \cdot \hat{m}_t \tag{3}$$

where θ_t is the model parameter at time step t, $\hat{m}_t$ and $\hat{v}_t$ are the bias-corrected estimates of the first and second moments, and ϵ is a small constant for numerical stability.

The categorical crossentropy [26] loss function was used to optimize the model, defined as:

$$\mathcal{L}_{\text{CCE}} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{4}$$

where C is the number of classes, y_i is the true label, and $\hat{y}_i$ is the predicted probability for class i.

To prevent overfitting and reduce unnecessary computation, EarlyStopping technique was applied during training. This monitors the validation loss and restores the best weights when no improvement is observed after 5 consecutive epochs. Figure 2 presents the architecture of the proposed model.
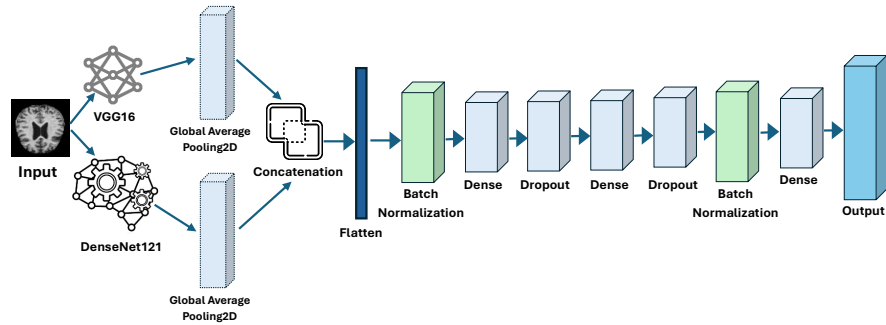

Fig. 2: Architecture of the proposed model

3.5   Model Evaluation Metrics

To evaluate the model's performance, several commonly used classification metrics were applied. These include accuracy, precision, recall, and F1-score. Each metric provides insight into different aspects of model performance.

Accuracy Accuracy measures the proportion of total correct predictions made by the model:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

Precision Precision shows how many of the predicted positive samples were actually correct:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6}$$

Recall Recall (or sensitivity) measures how many actual positive samples were correctly identified:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

F1-score F1-score is the harmonic mean of precision and recall. It provides a balanced metric, especially useful when classes are imbalanced:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

Here, TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

3.6   Explainable AI (XAI)

To better understand and interpret the decisions made by the proposed hybrid model, Explainable AI (XAI) [6] techniques were utilized. These techniques help highlight the regions of the input image that influenced the model's predictions. In this study, Grad-CAM and Integrated Gradients were used to generate visual explanations.

Grad-CAM Gradient-weighted Class Activation Mapping (Grad-CAM) [21] is a widely used technique to visualize the important regions in an image that contribute to a model's decision. Grad-CAM works by computing the gradient of the predicted class score with respect to the feature maps of the final convolutional layer. These gradients are then used to create a coarse heatmap that highlights discriminative regions.

The Grad-CAM heatmap $L^c_{\text{Grad-CAM}}$ for class c is computed as:

$$L^c_{\text{Grad-CAM}} = \text{ReLU}\left(\sum_k \alpha^c_k A^k\right) \tag{9}$$

where A^k represents the k-th feature map, and α^c_k is the weight computed by global average pooling over gradients:

$$\alpha^c_k = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A^k_{ij}} \tag{10}$$

Here, y^c is the score for class c, and Z is the number of pixels in the feature map. The ReLU function ensures only positive influences are visualized.

Integrated Gradients Integrated Gradients[27] was also employed to gain finer, pixel-level insights into the model's decisions. This method attributes the prediction of a model to its input features by integrating the gradients of the output with respect to the input along a straight path from a baseline (usually a black image) to the actual input.

Formally, the integrated gradient for the i-th input feature is given by:

$$\text{IG}_i(x) = (x_i - x'_i) \times \int_{\alpha=0}^{1} \frac{\partial F(x' + \alpha(x - x'))}{\partial x_i} d\alpha \tag{11}$$

where x is the input, x' is the baseline, and F is the model's output function. This approach helps understand which pixels in the image most significantly contributed to the prediction.

The combination of Grad-CAM and Integrated Gradients provides both coarse and fine-grained visual explanations, enhancing model transparency and building trust in clinical applications.

4   Results and Discussion

The performance of the proposed DGG-XNet model was evaluated on a multiclass brain MRI classification task involving three classes: Tumour, Normal, and Alzheimer's. The model achieved a test accuracy of 91.33%, significantly outperforming several well-established deep learning architectures.

Table 3 presents a comparison between DGG-XNet and other popular CNN-based models, including VGG16, DenseNet121, MobileNetV2, InceptionV3, ResNet

variants, and EfficientNetB3. Among these, VGG16 was the closest in performance, achieving an accuracy of 84.67%. Other models showed moderate to lower performance, with ResNet101 recording the lowest at 68.00%.

The strong performance of DGG-XNet can be attributed to the hybrid fusion of VGG16 and DenseNet121 feature extractors, enabling the model to capture both local and global patterns effectively. Additionally, techniques such as transfer learning, fine-tuning, data balancing, and the use of early stopping helped improve generalization while avoiding overfitting.

Table 3: Model Accuracy Comparison

Model	Accuracy
DGG-XNet (proposed model)	**91.33%**
VGG16	84.67%
DenseNet121	82.67%
InceptionV3	82.00%
MobileNetV2	81.33%
ResNet50	78.00%
EfficientNetB3	74.67%
ResNet101	68.00%

The results confirm that DGG-XNet offers a more accurate and reliable solution for brain MRI classification, and demonstrates the effectiveness of hybrid feature fusion in complex medical imaging tasks.

4.1 Training and Validation Performance

Figure 3 shows the training and validation accuracy and loss curves for the proposed model. The curves indicate that the model was able to learn effectively, with both training and validation accuracy gradually increasing and validation loss stabilizing, suggesting minimal overfitting.

4.2 ROC Curve Analysis

To evaluate the model's ability to distinguish between the classes, ROC curves were plotted for each class using one-vs-rest classification. As shown in Figure 4b, the proposed model achieved high AUC scores for all three classes, confirming its strong classification performance.
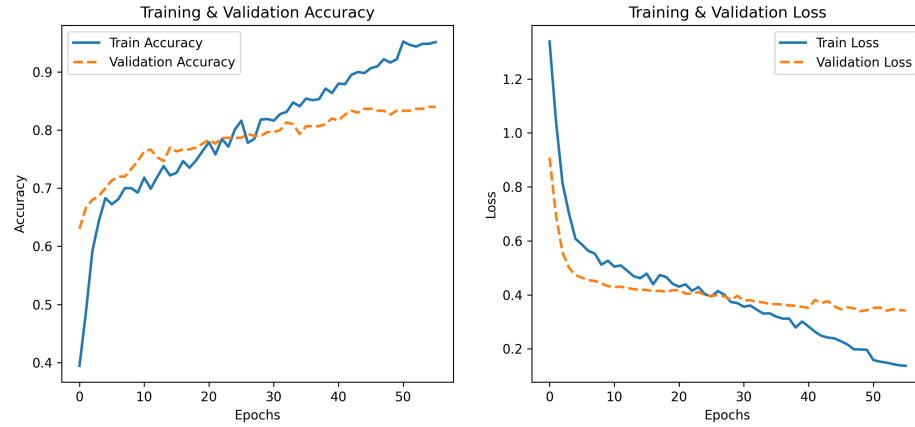
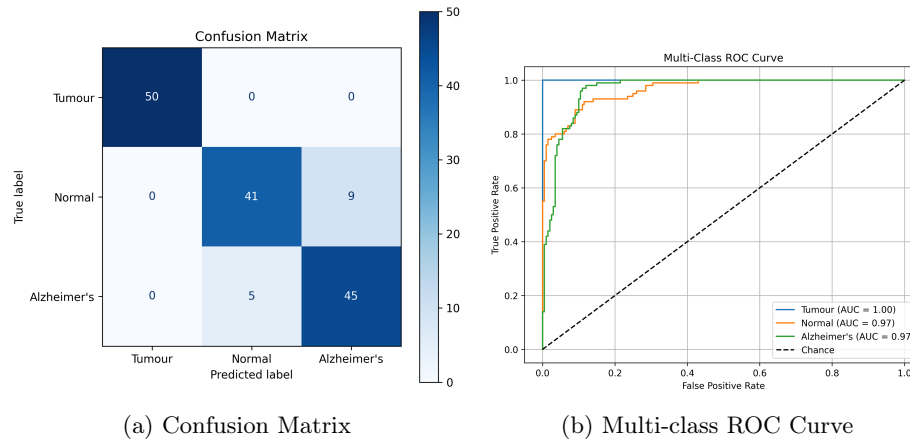Fig. 3: Training and validation accuracy and loss curves.

(a) Confusion Matrix

(b) Multi-class ROC Curve

Fig. 4: Performance evaluation: Confusion Matrix and ROC Curve

4.3   Confusion Matrix

Figure 4a illustrates the confusion matrix based on a balanced test set (50 samples per class). The model demonstrated a high number of correct predictions across all classes, with very few misclassifications, further validating its effectiveness.

4.4   Explainable AI Visualizations

To gain insight into the decision-making process of the proposed model, explainable AI techniques such as Grad-CAM and Integrated Gradients were applied. Figure 5 displays these visualizations. Grad-CAM highlights the important regions of the image that influenced the model's prediction, while Integrated Gradients provides pixel-level attribution. Together, these methods enhance the interpretability and transparency of the model, making it more trustworthy in clinical scenarios.
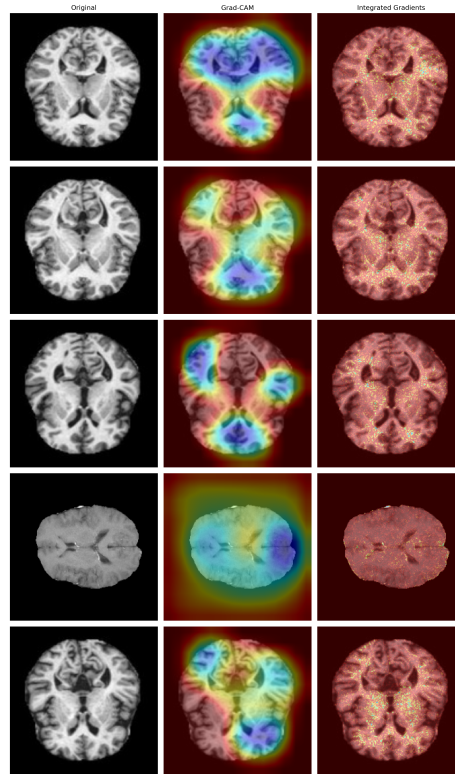

Fig. 5: Explainable AI: Grad-CAM & Integrated Gradients

4.5   Evaluation Metrics

In addition to accuracy, precision, recall, and F1-score were computed to provide a more comprehensive evaluation of the proposed model. The results are summarized in Table 4, based on a test set of 150 samples. The model achieved an overall accuracy of 91.33%, with a macro-average F1-score of 90.00%.

Table 4: Precision, Recall, and F1-score of the proposed model on the test set.

Class	Precision	Recall	F1-Score
Tumour	1.00	1.00	1.00
Normal	0.89	0.87	0.88
Alzheimer's	0.82	0.84	0.83
Macro Avg	0.90	0.90	0.90
Accuracy	91.33%		

5   Conclusion

This study proposed a hybrid deep learning model, DGG-XNet which was designed for multi-class classification of brain MRI images into Tumour, Normal, and Alzheimer's categories. The architecture effectively combines the feature extraction strengths of VGG16 and DenseNet121, supported by global average pooling, feature fusion, and dense classification layers. Techniques such as transfer learning, class balancing, and fine-tuning were incorporated to improve generalization and robustness. The model achieved a high test accuracy of 91.33%, outperforming several established deep learning architectures. Quantitative evaluation using precision, recall, and F1-score confirmed consistent performance across all classes. Furthermore, explainable AI techniques such as Grad-CAM and Integrated Gradients provided valuable visual insights into the model's decision-making process, enhancing its transparency and trustworthiness in clinical contexts.

Future work may explore integrating 3D volumetric analysis, additional imaging modalities, Data augmentation techniques, and domain adaptation to improve performance across varied datasets and real-world scenarios.

References

1. Agarap, A.F.: Deep learning using rectified linear units (relu). ArXiv (2018)
2. Ayon, S.S., Hossain, M.E., Miah, M.S.U., Rahman, M.M., Mahmud, M.: Advancing mental health problems with machine learning and genetic algorithms for anxiety classification in bangladeshi university students. In: Itthipuripat, S., Ascoli, G.A., Li, A., Pat, N., Kuai, H. (eds.) Brain Informatics. pp. 338–350. Springer Nature Singapore, Singapore (2025)

3. Baid, U., et al.: The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv:2107.02314 (2021)
4. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., et al.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. Nature Scientific Data **4**, 170117 (2017)
5. Castellazzi, G., Cuzzoni, M.G., Cotta Ramusino, M., Martinelli, D., Denaro, F., Ricciardi, A., Vitali, P., Anzalone, N., Bernini, S., Palesi, F., Sinforiani, E., Costa, A., Micieli, G., D'Angelo, E., Magenes, G., Gandini Wheeler-Kingshott, C.A.M.: A machine learning approach for the differential diagnosis of alzheimer and vascular dementia fed by mri selected features. Frontiers in Neuroinformatics **14** (2020)
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 248–255 (2009)
7. Ghazal, T., Abbas, S., Munir, S., Khan, M., Ahmad, M., Issa, G., Zahra, S., Khan, M., Hasan, M.: Alzheimer disease detection empowered with transfer learning. Computers, Materials and Continua **70**(3), 5005–5019 (2022)
8. Ghnemat, R., Khalil, A., Abu Al-Haija, Q.: Ischemic stroke lesion segmentation using mutation model and generative adversarial network. Electronics **12**(3) (2023)
9. Greutter, L., Miller-Michlits, Y., Klotz, S., Reimann, R., Nenning, K.H., Platzek, S., Krause, E., Kiesel, B., Widhalm, G., Langs, G., Baumann, B., Woehrer, A.: Frequent alzheimer's disease neuropathological change in patients with glioblastoma. Neuro-Oncology Advances **6**(1), vdae118 (2024)
10. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2261–2269 (2017)
11. Huang, J., et al.: Altered d-glucose in brain parenchyma and cerebrospinal fluid of early alzheimer's disease detected by dynamic glucose-enhanced mri. Science Advances **6**(eaba3884) (2020)
12. Jahn, H.: Memory loss in alzheimer's disease. Dialogues in Clinical Neuroscience **15**(4), 445–454 (2013)
13. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. ArXiv (2014)
14. Lerousseau, M., Deutsh, E., Paragios, N.: Multimodal brain tumor classification. arXiv preprint arXiv:2009.01592 (2020)
15. Lin, M., Chen, Q., Yan, S.: Network in network. ArXiv (2013)
16. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE Transactions on Medical Imaging **34**(10), 1993–2024 (2015)
17. Narayana, T.L., Reddy, T.S.: An efficient optimization technique to detect brain tumor from mri images. In: 2018 International Conference on Smart Systems and Inventive Technology (ICSSIT). pp. 168–171 (2018)
18. Poma, Y., Melin, P., González, C.I., Martínez, G.E.: Optimization of convolutional neural networks using the fuzzy gravitational search algorithm. Journal of Automation, Mobile Robotics and Intelligent Systems **14**(1) (2020)
19. Raghuram, S., Bharadwaj, A.S., K, D.S., Khadabadi, M.S., Jayaprakash, A.: Digital implementation of the softmax activation function and the inverse softmax function. In: 2022 4th International Conference on Circuits, Control, Communication and Computing (I4C). pp. 64–67 (2022)
20. Rammurthy, D., Mahesh, P.: Whho-based deepcnn for brain tumor detection using mri images. In: 2022 IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3259–3272 (2022)

21. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: 2017 IEEE International Conference on Computer Vision (ICCV). pp. 618–626 (2017)
22. Sharma, S., Guleria, K., Tiwari, S., Kumar, S.: A deep learning based convolutional neural network model with vgg16 feature extractor for the detection of alzheimer disease using mri scans. Measurement: Sensors **24**, 100506 (2022)
23. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: Proceedings of the International Conference on Learning Representations (ICLR). San Diego, USA (2015)
24. World Health Organization: Over 1 in 3 people affected by neurological conditions: The leading cause of illness and disability worldwide (2024)
25. Zaw, H.T., Maneerat, N., Win, K.Y.: Brain tumor detection based on naïve bayes classification. In: 2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST). pp. 1–4 (2019)
26. Zhang, Z., Sabuncu, M.R.: Generalized cross entropy loss for training deep neural networks with noisy labels. ArXiv (2018)
27. Zhuo, Y., Ge, Z.: Ig2: Integrated gradient on iterative gradient path for feature attribution. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(11), 7173–7190 (2024)