

SvedbergOpen
DISSEMINATION OF KNOWLEDGEInternational Journal of Artificial
Intelligence and Machine LearningPublisher's Home Page: <https://www.svedbergopen.com/>

Research Paper

Open Access

Vision Transformer-Based Framework for Multi-Class Neurological Disease Detection Using Medical Imaging

Aruna Kokkula^{1*}, Venkataramana Battula², Radhika Kondam³, P. Chandra Sekhar⁴, B. Kranthi Kiran⁵¹Sr. Assistant Professor, Department of ECE, MVSR Engineering College, Hyderabad, India, E-mail: arunak_ece@mvsrec.edu.in
ORCID ID: <https://orcid.org/0000-0002-5697-2226>²Sr. Assistant Professor, CSE-Allied(AI&ML) Department, MVSR Engineering College, Hyderabad, India,
E-mail: venkataramana_cse@mvsrec.edu.in; ORCID ID: <https://orcid.org/0000-0002-2548-3964>³Director of Operations, Cognisys AI Pvt. Ltd., Hyderabad, India, Email: radhikakondam@gmail.com⁴Professor, Department of ECE, University College of Engineering, Osmania University, Hyderabad, India, E-mail: sekharp@osmania.ac.in,
ORCID ID: <https://orcid.org/0000-0003-0228-0571>⁵Professor, CSE Department, JNTU Hyderabad (JNTUH), Hyderabad, India, E-mail: kranthi@jntuh.ac.in, ORCID: <https://orcid.org/0009-0007-4141-2182>**ABSTRACT**

Neurological illnesses like Alzheimer's disease, brain tumors, multiple sclerosis, and other neurodegenerative disorders need precise and early diagnosis to be effectively treated clinically and to manage the disease. Classical diagnostic methods are often lengthy and heavily reliant on specialist interpretation of the clinical picture. To overcome these issues, this paper proposes a Vision Transformer-based design for detecting multi-class neurological diseases with state-of-the-art deep learning methods. The proposed framework applies the pretrained vit_base_patch16_224 architecture, utilized to extract global, contextual, and spatial features on neurological medical images. Preprocessing of the MCND dataset involves image resizing, normalization, and transformation of tensors before training a model. Transfer learning is used to enhance prediction and predictive efficiency. Evaluation of the framework using various performance metrics is assessed in terms of accuracy, precision, recall, F1-score, confusion Matrix analysis, and ROC-AUC curves. The experimental findings show that the proposed model has stable and high classification performance on various neurological disease categories with successful learning behavior and low overfitting. The framework also exhibits high robustness, scalability, and the ability to render consistent predictive capabilities in intelligent analysis of medical images. It offers a growing foundation for the subsequent incorporation of Explainable AI and progressive deep learning systems in neurological health care systems.

Keywords: Deep Learning, Medical Image Analysis, Multi-Class Classification, Neurological Disease Detection, Transfer Learning, Vision Transformer.

Introduction

There are neurological diseases like Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, brain tumours, and other rapidly increasing diseases around the globe, and they have become significant healthcare issues, as they have a devastating effect on human health and quality of life. Timely and correct diagnosis of these disorders is crucial for effective treatment planning, monitoring of the disease, and better patient outcomes. Traditional methods of diagnosis mainly depend on clinical examination and manual analysis of neuroimaging like MRI, CT, PET, and EEG. Nevertheless, such conventional techniques can be time-consuming, expensive, and highly reliant on expert judgment, thus prone to delayed diagnosis and poor diagnostic performance.

New developments in Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) have contributed tremendously to the enhancement of intelligent healthcare systems and medical image analysis. Convolutional Neural Networks (CNNs) and other deep learning models have been shown to be able to extract meaningful features by automatically analysing medical images to perform disease classification tasks. CNN-based methods have shown good performance; however, they are frequently unable to model long-range dependencies and contextual interactions of globally aligned image data in complex neurological images. These constraints minimize their capability to specify various categories of neurological conditions.

Transformer-based architectures have now become a significant topic of medical image analysis in order to address these challenges. ViTs subdivide input images into smaller patches and process them with a mechanism of self-attention that effectively learns the global contextual details. Vision Transformers are also better models when compared to conventional CNN models, have better features, are more scalable, and perform better at complex neuroimaging applications. Pretrained transformer models also perform better at transfer learning, which improves computational efficiency and speeds up training.

The proposed work presents a Vision Transformer scheme for multi-class automatic detection of neurological diseases in medical images. The architecture involves the vitbase_patch16_224 model, which is pretrained to

extract high-level features such as spatial and contextual details, within the context of neurological medical images. The MCND neurological disease dataset is obtained via the KaggleHub platform and processed with image resizing, image normalization, image tensor transformations, and data augmentation to make the information consistent with an efficient learning process. Multi-class classification of neurological disorders with the processed images is then performed. The proposed framework uses the Adam optimizer and the Cross-Entropy Loss function to optimize the model efficiently and converge the model in a stable manner. Multiple performance measures are used to assess the performance of the proposed framework, such as accuracy, precision, recall, F1-score, confusion matrix analysis, and ROC-AUC curves. Experimental evidence has shown that the suggested Vision Transformer-based model has good classification accuracy, stable convergence, good generalization, and reliable disease prediction in various types of neurological diseases. Generally, the presented framework is a scalable and efficient method of intelligent diagnosis of neurological diseases with medical imaging data.

The proposed work's main contribution is the classification of different neurological diseases from medical images using a pre-trained ViT-Base-Patch16-224 model. The model is used to learn features from the images, both global and local. The input data is enhanced using image pre-processing methods such as resizing, normalization, tensor conversion, and data augmentation. We use transfer learning to improve feature extraction and classification performance as well as reduce computation. The proposed model is evaluated using accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC. The results obtained show that the proposed method has good and stable performance in the identification of various neurological diseases.

2. Literature Review

Effective treatment and management of the disease require early and accurate diagnosis of neurological diseases like Alzheimer's, brain tumours, Parkinson's, multiple sclerosis, as well as other neurodegenerative diseases. Medical imaging, especially Magnetic Resonance Imaging (MRI), is an important procedure that is used in the diagnosis of neurological diseases. Nevertheless, conventional approaches to diagnosis are sometimes lengthy and highly reliant on medical skills. The current development of deep learning, particularly the application of Vision Transformers (ViTs) and hybrid CNN-transformer models, has seen recent research on automated neurological disease detection and classification with remarkable improvement.

The hierarchical multi-scale Vision Transformer model that was proposed by Sankari et al. [1] was aimed at detecting and classifying brain tumours with the help of MRI images. Their framework was able to effectively describe local and global contextual features, by which the multi-class classification performance was enhanced. Likewise, Reddy et al. [2] proposed a fine-tuned Vision Transformer-based model to improve the classification of brain tumours by using MRI scan images. They proved to be better in terms of feature extraction and better in classification than more classic CNN-based techniques. In the case of the diagnosis of Alzheimer's disease, Almansour et al. [3] suggested a deformable attention Vision Transformer model that was capable of recording spatial relationships on the MRI scans. Joy et al. [5] designed ViTAD, an adapted version of the Vision Transformer architecture, to perform the task of multi-stage classification of Alzheimer's disease with high performance on the disorders at various stages. Ahmed [25] also developed a Dual-stage Vision Transformer with attention-guided patch aggregation, which improved the detection of Alzheimer's disease to a high level by learning better features.

There has also been a lot of interest in hybrid deep learning models that combine CNNs with Vision Transformers. Tariq et al. [4] added Vision Transformers to EfficientNetV2 to classify brain tumors into multiple classes, enhancing the feature representations and classification performance. The framework of explainable medical image classification presented by Hussein et al. [10] is called EFFResNet-ViT and is a combination of CNN and Vision Transformer that uses a fusion model. Their method increased interpretability without compromising classification performance. For multi-class brain tumor classification and computer-aided intervention assistance, Hybrid-CNNViT, proposed by Chowdhury et al. [18], showed the success of hybrid architectures in the context of medical imaging. Multimodal image analysis in medical imaging has been proposed as a significant area of research towards enhancing diagnostic accuracy. Kaliappan et al. [6] introduced a single Vision Transformer framework for multimodal medical image classification with a focus on the advantages of the combination of various imaging modalities. Fasnamol et al. [14] created an explainable cross-attention hybrid CNN-transformer network for the diagnosis of multimodal brain diseases. Their model enhanced the fusion and interpretability of multimodal features by virtue of cross-attention. Sameera et al. [11] also indicated that transformers are increasingly becoming important in multimodal healthcare image analysis. Several papers revolved around the systems of attention and transformer architectures for classifying neurodegenerative diseases. Saha et al. [7] released AGHC-NeuroTransformer, an attention-based hybrid CNN-transformer, to achieve strong classification of neurodegenerative disorders in MRI scans. In their study, MR and Kavitha [16] employed Swin Transformers to classify multi-class neurodegenerative diseases based on brain MRI images. Gunver [21] explored the use of deep learning models using transformers to detect Alzheimer's and Parkinson's disease and found that the transformer architecture was effective in the analysis of neurological diseases.

Explainable AI (XAI) is ever gaining relevance in smart healthcare systems to enhance the transparency of models and clinical trust. Rahman et al. [8] developed an explainable multi-class eye disease-detecting cross-attention fusion Vision Transformer. Panigrahi et al. [24] proposed a framework of ViT-B16 transformer-based interpretable brain MRI tumor classification. Imran et al. [20] suggested NeuroViT, which is a neuro-adaptive predictive coding Vision Transformer that classifies ocular disease in a way that can be easily interpreted.

My fellow researchers have also tested ensemble learning and feature fusion strategies to enhance the classification performance. Chakra Bortty et al. [12] developed a streamlined set of Vision Transformers and CNNs to detect the presence of Alzheimer's in medical images. Konatham et al. [23] employed multi-stage feature fusion to closely combine Vision Transformers, residual CNNs, and channel attention to improve brain tumor detection. In the classification of eye disease, AlMohimeed [19] has suggested a multi-stage framework that is based on transformers with feature fusion and feature learning.

Other recent research has been on novel transformer architectures and effective learning schemes. Priyan et al. [17] suggested a thin slicing framework of transformers for exact tumor location and in vivo stage evaluation of ultra-high-resolution medical imaging. Ali et al. [26] presented the X-ViT-CNN, a combination of transfer learning and customized Vision Transformers to predict the Alzheimer's disease of the patient in several stages based on the MRI images. Jeyaprabha et al. [27] proposed a dual-attention-enhanced framework of transformers to diagnose Alzheimer's in multi-classes by utilizing brain MRI images.

Galvanic studies also focus on the shift towards Vision Transformer-based approaches rather than on traditional CNNs. Aamir et al. [28] provided an extensive review of the dynamic developments in deep learning architectures in brain tumor classification, where transformer models have been on the rise. Alrahhal et al. [29] employed Data-Efficient Vision Transformers (DEiT) that are used in the high-precision multi-class diagnosis of brain tumors. Gomez-Guzmana et al. [30] utilized CNNs and transformer architectures together (pre-trained) on the classified images of brain tumors with the use of MRI. Velu and Jaisankar [31] came up with a CNN-Swin Transformer hybrid architecture for foreseeing Alzheimer's disease and indicated the benefits of integrating convolutional-based feature extraction with hierarchical-based learning in transformers.

Though Vision Transformers and CNN-based models have been extensively used to detect neurological diseases in the literature, most models consider individual disease types, including Alzheimer's disease or brain tumors [1], [3], [5]. Several hybrid CNN-transformer models also have the disadvantages of being computationally complex and incapable of successfully representing the global contextual content of medical images [10], [18]. Moreover, current approaches can usually be characterized by the difficulties of scalability, resilience, ambivalent interpretability, paucity of annotated medical datasets, and deficient generalization skills in terms of multi-class neurological disease classification [21], [28]. Recent reports indicate that Vision Transformers and hybrid architectures are crucial in increasing the detection of neurological diseases due to their superior capabilities in feature extraction, self-attention, multimodal fusion presentation, and explainable learning. Nevertheless, the challenge of building a strong, scalable, and efficient model to monitor the automated identification of multiple neurological diseases is a significant topic to be examined in research. The proposed ViT framework can be used to overcome these drawbacks to accurately, scalably, and effectively diagnose multi-class neurological diseases.

2.1 Proposed Work

The work suggests a structure based on a Vision Transformer (ViT) model to detect and classify neurological diseases in medical images as a multi-class object. The system entails preprocessing, feature extraction, classification, and evaluation. Transfer learning enhances performance and efficiency. The model is trained with the Adam optimizer and tested with accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC.

2.2 System Architecture

The proposed system architecture has six steps, which include acquisition of the dataset, preprocessing, feature extraction through Vision Transformer, classification, training of the model, and evaluation of performance. The MCND data is gathered on the KaggleHub site, and it is categorized into: Alzheimer's disease, brain tumors, multiple sclerosis, Parkinson's disease, and normal brain conditions. In preprocessing, the images will be resized to 224 x 224, normalized, turned into tensors, and augmented with random flipping, rotation, and changing brightness. The data is split into training and testing sets in the proportion of 80:20.

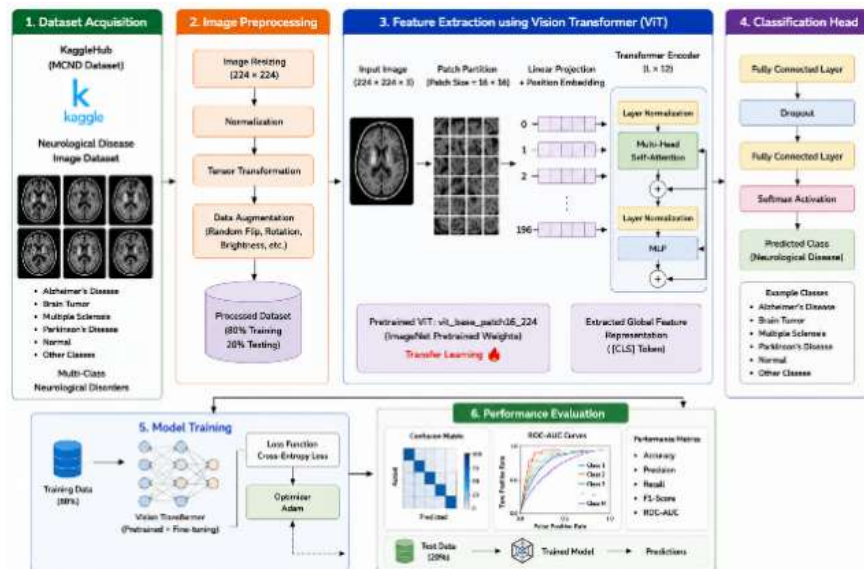


Figure 1: System architecture

The input to the pretrained ViT-base-patch16-224 model is the preprocessed images. The Vision Transformer separates the images into fixed-size patches and applies the multi-head self-attention and multilayer perceptron (MLP) layers to learn the contextual and spatial features on a global scale. The features extracted are subjected to fully connected layers and softmax activation to classify them. The framework is based on the Cross-Entropy Loss and Adam optimizer to efficiently train and stabilize. Lastly, the model is analyzed on the basis of accuracy, precision, recall, F1-score, confusion analysis, and ROC-AUC curves.

3. Methodology

The data analysis paradigm is aimed at adopting a deep learning model, DeepVision Transformer, to classify neurological diseases. Initially, the MCND dataset is separated into training and testing sets in an 80:20 proportion. To minimize overfitting and enhance data consistency, preprocessing and augmentation methods are used. The second step is to use transfer learning with the trained vit_base patch16 permit Model of the ability to extract deep features by utilizing the previously trained vit base patch16 224 model on the given medical images in the neurology field. A transformer encoder is trained to use patch embeddings and multi-head self-attention to learn discriminative representations. The learned features fall into several disease categories with fully connected and softmax layers. Backpropagation during training chooses the parameters of the network with the Adam optimizer. Lastly, the trained model is evaluated on unseen samples, and its performance is gauged in terms of accuracy, precision, recall, F1-score, confusion matrix analysis, and ROC-AUC evaluation metrics.

Equation No.	Equation	Description	
1	Self-Attention Mechanism	$Attention(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$	Computes relationships between image patches using query (Q), key (K), and value (V) matrices. The scaling factor d_k stabilizes the attention weights. It enables the model to capture long-range spatial dependencies and global contextual information.
2	Image Patch Embedding	$Z_0 = [x_p^1 E; x_p^2 E; \dots; x_p^N E] + E_{pos}$	Divides the input image into N fixed-size patches. Each patch x_p^i is projected into an embedding vector using linear projection matrix E . Positional encoding E_{pos} is added to preserve spatial information.
3	Cross-Entropy Loss Function	$L = -\sum_{i=1}^N y_i \log(\hat{y}_i)$	Measures the difference between true labels y_i and predicted probabilities $\hat{y}_i$. It minimizes classification errors during training and improves the prediction capability of the model.
4	Adam Optimizer Update Rule	$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}}$	Updates model parameters θ_t at iteration t using learning rate η . $\hat{m}_t$ and $\hat{v}_t$ are bias-corrected first and second moment estimates, and ϵ is a small constant for numerical stability.
5	Classification Accuracy	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$	Evaluates overall prediction performance. TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative) are used to compute the proportion of correctly classified samples.
6	Precision	$Precision = \frac{TP}{TP + FP}$	Measures the correctness of positive predictions. It indicates the proportion of predicted positive samples that are actually correct.
7	Recall	$Recall = \frac{TP}{TP + FN}$	Measures the ability of the model to identify actual positive samples. It indicates the proportion of actual positives that are correctly detected by the model.
8	F1-Score	$F1-Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$	Represents the harmonic mean of precision and recall. It provides a balanced evaluation metric, especially useful for imbalanced datasets in multi-class neurological disease classification.

Figure 2: Mathematical Formulations and Performance Evaluation Metrics of the Proposed Vision Transformer-Based Neurological Disease Detection Framework

The algorithm introduces a Vision Transformer-based approach to automated multi-class neurological disease detection with MRI images, including preprocessing, patch embedding, self-attention-based feature extraction,

transfer learning, optimization, and performance evaluation to enable accurate, scalable neurological disease classification.

Table 1: Algorithm of Proposed Vision Transformer-Based Framework for Multi-Class Neurological Disease Detection

Input: MCND Brain MRI Dataset (D) Output: Predicted Neurological Disease Class (C) Dataset Acquisition: Load the MCND dataset containing MRI images from multiple neurological disease categories. MRI Image Preprocessing: Resize all MRI images to (224 X 224) and apply normalization and tensor transformations. Dataset Partitioning: Divide the dataset into a training set ($D_{\{train\}}$) and a test set ($D_{\{test\}}$). Data Augmentation: Apply augmentation techniques such as rotation, flipping, and transformation to improve model generalization. Model Initialization: Initialize the pretrained Vision Transformer model vit_base_patch16_224. Patch Generation and Embedding: Divide each MRI image into fixed-size patches and generate embedding vectors with positional encoding using Equation (2). Self-Attention Feature Learning: Pass embedded patches through transformer encoder layers using the self-attention mechanism defined in Equation (1). Global Feature Extraction: Extract global contextual and spatial features from neurological MRI images. Multi-Class Classification: Apply a fully connected classification layer for predicting neurological diseases. Loss Computation: Compute Cross-Entropy Loss using Equation (3) between predicted labels and ground truth labels. Parameter Optimization: Update model parameters using the Adam optimizer defined in Equation (4). Iterative Transformer Training: Repeat Steps 6–11 for multiple epochs until the convergence criteria are satisfied. Performance Evaluation: Evaluate the trained model using Accuracy, Precision, Recall, and F1-Score metrics computed using Equations (5), (6), (7), and (8), respectively. Neurological Disease Inference: Predict the neurological disease category for unseen MRI images.

3.1. Implementation

The suggested architecture is executed with Python and the PyTorch deep learning framework on a Google Colab server. The dataset of neurological diseases is downloaded with the help of the KaggleHub API and loaded with the ImageFolder utility in torchvision. Image pre-processing tasks involve resizing the input images to (224 X 224) pixels, normalization, and transforming the images into tensors.

Timm library and the Vision Transformer model are used to implement the work and use the pre-trained ImageNet weights to enhance the transfer learning process. The last classification level is adjusted depending on the number of neurological disease types within the dataset. An adaptive learning rate is used with the Adam optimizer to optimize convergence in the network. It is trained using 10 epochs with computational efficiency through the use of GPU acceleration.

Performance visualization modules to create confusion matrices, ROC curves, and training-validation accuracy plots are also implemented in the implementation. The structure computes weighted precision, recall, and F1 - score with Scikit-learn assessment tools.

3.2. Dataset

Multi-Class Neurological Disorder[32] dataset is a publicly accessible dataset of medical imaging that includes brain MRI scans of a variety of neurological disorders. The dataset is a combination of the MRI images of Alzheimer's Disease (AD), Brain Tumor (BT), and Multiple Sclerosis (MS) datasets on the Kaggle repository. The MCND dataset includes 16,400 human brain MRI scans that are divided into eight classes of unknown cases: AD_MildDemented, AD_ModerateDemented, AD_VeryMildDemented, BT_glioma, BT_meningioma, BT_pituitary, MS, and Normal (healthy). The dataset is popular in the field of multi-class classification of neural diseases and in analyzing medical images through deep learning and Vision Transformer systems.

3.3. Experimental Setup

The experimental system is a hardware and software configuration of the proposed framework training and evaluation. The experiments were done in the Google environment using the help of GPUs to speed up the computation of transformer-based deep learning. Python was used, and some libraries were used, including PyTorch, TorchVision, TIMM, NumPy, Matplotlib, Seaborn, and Scikit-learn. Experimentation was done on the MCND dataset, which comprises several neurological disease classes. The data was divided into two parts, training and test sets, in the ratio of 80: 20. Data loading was done in batches based on DataLoader utilities to maximise the use of memory and computational power. The Vision Transformer model vit_base_patch16_224 was initialised with ImageNet pretrained weights and was trained for 10 epochs with the Adam optimizer and Cross-Entropy Loss for optimisation using efficient gradient-based optimisation. Table 3 shows the hyperparameters and the training parameters of the suggested Vision Transformer-based model.

Table 3: Hyperparameter Settings and Training Configuration of the Proposed Vision Transformer-Based Framework

Parameter	Value
Model	vit_base_patch16_224
Optimizer	Adam
Epochs	10
Batch Size	32
Learning Rate	0.0001
Input Size	224×224
Loss Function	Cross-Entropy
Train-Test Split	80:20

4. Results

The hypothesis test of the suggested Vision Transformer (ViT)-based architecture proves its efficiency in terms of automated detection of neurological diseases in medical samples across multiple classes using medical images. The advocated framework was able to acquire discriminative spatial and global contextual features from neurological MRI images with high classification accuracies across various disease categories. The experimental outcome validates that transformer-based architectures are very appropriate in tasks involving rich neuron-based image analysis.

Table 4 displays the ablation experiment done in various experimental conditions. The findings show that preprocessing methods, transfer learning, and transformer-based feature extraction greatly enhanced the performance of the neurological disease classification. The CNN model without preprocessing had an accuracy of 84.7, and with preprocessing, the CNN had an accuracy of 87.9. Likewise, the Vision Transformer model without transfer learning reached 91.2 percent accuracy, but the accuracy rose to 92.8 percent in the case of preprocessing. The suggested Vision Transformer framework (this time with transfer learning) achieved a classification rate of 94.1%, showing that transfer learning and extractions of global contextual features are indeed good for automatic disease diagnosis in neurology.

Table 4: Ablation Study of the Proposed Vision Transformer-Based Framework under Different Experimental Configurations

Configuration	Accuracy
CNN without preprocessing	84.7%
CNN with preprocessing	87.9%
ViT without transfer learning	91.2%
ViT with preprocessing	92.8%
Proposed ViT + Transfer Learning	94.1%

Generally, the experimental findings support the efficacy of the suggested Vision Transformer-based structure in automated multi-class labeling of neurological diseases with the use of MRI images. Transformer-based feature extraction, preprocessing, and transfer learning proved to be excellent in enhancing the prediction performance over traditional CNN-based methods. The suggested architecture has good promise for intelligent healthcare systems, automated diagnosis of neurological diseases, and integration with Explainable AI and advanced deep learning algorithms in the future.

Figure 3 shows the trend of training and validation accuracy in the various epochs. The training accuracy was raised to a higher value of 94.1, and the validation accuracy was augmented to a higher value of 94.1 compared to 77.8 and 78.8, respectively. The good flow of the two curves depicts efficient learning behaviour, stable

convergence, and efficient feature extraction of the proposed structure. The fact that training and validation accuracies are nearly equal, which implies that there is little overfitting and effective generalization to unseen neurological images. The analysis of the confusion matrix of several classes of neurological diseases (including the stages of Alzheimer's disease, types of brain tumors, multiple sclerosis, and normal brain conditions) is provided in Figure 4. The majority of classes of neurological diseases were properly identified with minimal interclass confusion. The framework proposed was especially effective in classifying BT_glioma, BT_pituitary, Normal, and AD_VeryMildDemented. There was some minor misgrouping between related disease groups as a result of similarity between MRI image patterns. In general, the confusion matrix reflects the ability of the Vision Transformer-based feature extraction mechanism to be discriminative.

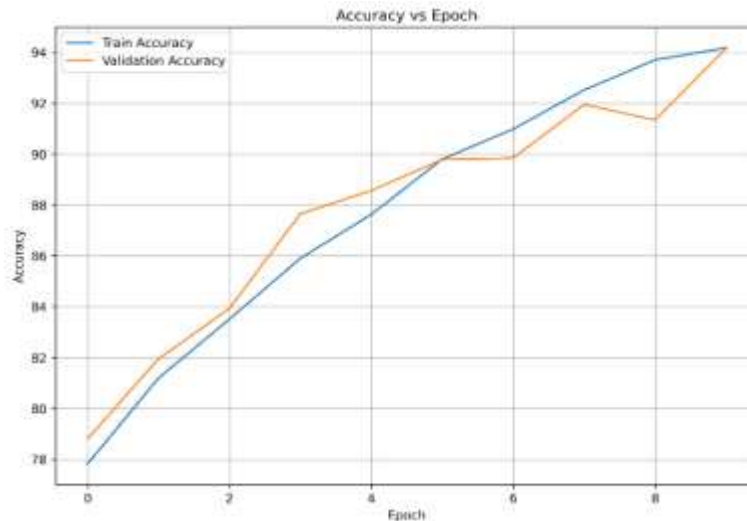


Figure 3: Accuracy Vs Epoch



Figure 4: Confusion Matrix

Table 6: Classification Performance of the Proposed Framework

Class	Precision	Recall	F1
AD_MildDemented	0.93	0.87	0.90
AD_ModerateDemented	1.00	1.00	1.00
AD_VeryMildDemented	0.89	0.88	0.89
BT_glioma	0.99	0.98	0.98
BT_meningioma	0.98	0.96	0.97
BT_pituitary	0.98	0.99	0.98

Class	Precision	Recall	F1
MS	0.99	0.83	0.90
Normal	0.93	0.97	0.95
Overall Accuracy			94%

Table 6 shows the specific classification in the proposed framework using the accuracy, recall, and F1-scores. The framework obtained a total classification rate of 94%. The BTglioma, BTpituitary, and AD moderate dementia categories exhibited especially high accuracy and recall, and showed consistent classifications across a variety of neurological disease categories. Moreover, the total run-time needed to train and evaluate the model is about 107.98 minutes, indicating that it is possible to run model-based applications of intelligent medical image analysis with reasonable computational efficiency.

Figure 5 shows the distribution of the dataset when experimenting. There are several categories of neurological diseases in the MCND dataset, each having a different sample size, where the normal type has the highest number of MRI images. The proposed framework, despite disproportionately combined classes, showed consistent predictive accuracy with both majority and minority classes, revealing strength and scalability to multi-class neurological disease identification.

Figure 6 shows the time complexity of the computations performed by the proposed framework per epoch. The time of execution was fairly steady during the training process, lying between 604 and 613 seconds on average time taken over all the epochs. The regularity of the computation performance shows good GPU usage and scalability of the Vision Transformer architecture to analyze a large-scale neurological disease based on this architecture.

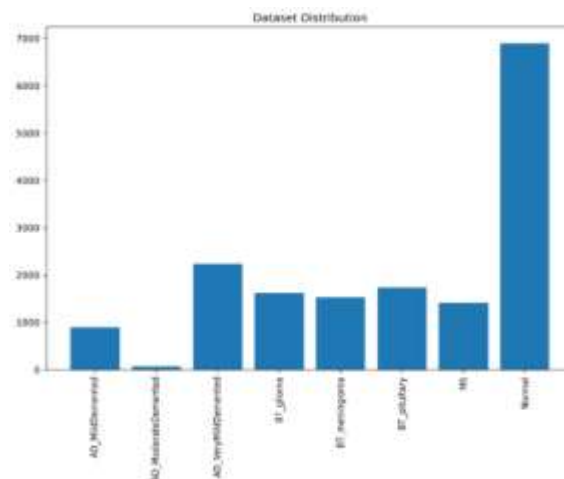


Figure 5: Dataset Distribution

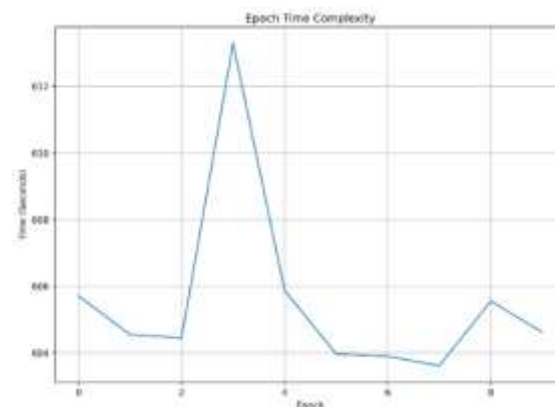


Figure 6: Epoch Time Complexity

Figure 7 shows the training and validation loss curves during model training. During training, the training loss dropped to 0.15, and the validation loss to 0.17. The constant decline of the loss statistics attests to steady convergence, effective Pearson optimization of the model parameters, and better learning ability of the suggested structure based on the Adam optimizer and Cross-Entropy Loss ratio.

The distribution of prediction that is obtained by the trained model is shown in Figure 8. The predicted distribution is very similar to the original dataset distribution, which means that the ability to make a balanced

classification is achieved over different neurological diseases. The framework showed similar predictive performance on categories of diseases that were trained using smaller numbers.

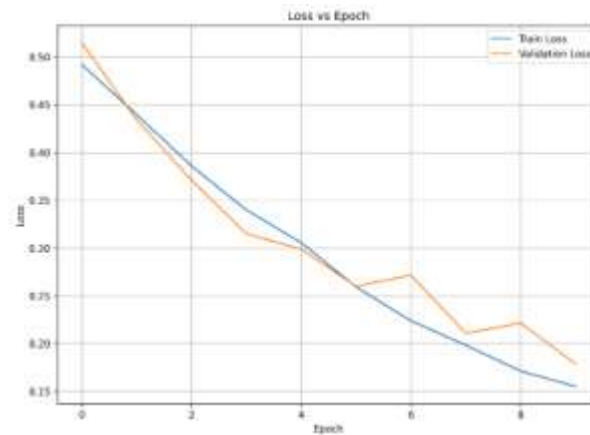


Figure 7: Loss Vs Epoch

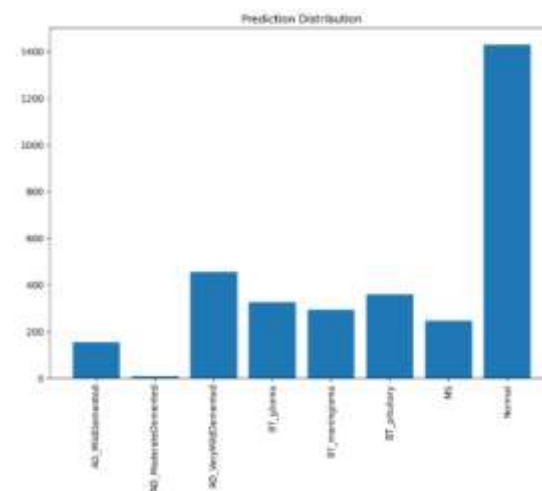


Figure 8: Prediction Distribution

The analysis of the Receiver Operating Characteristic (ROC) curve presented in Figure 9 illustrates the great discriminative ability of the suggested framework based on Vision Transformers. The model was found to be 0.99 to 1.00 in all the neurological disease categories in terms of Area Under the Curve (AUC). Especially, the near-perfect performance of the AUC scores for the disease classes AD_MildDemented, AD_ModerateDemented, BT_glioma, BT_meningioma, and BT_pituitary confirmed high separability of the disease classes and very good performance for the neurological diseases that were predicted.

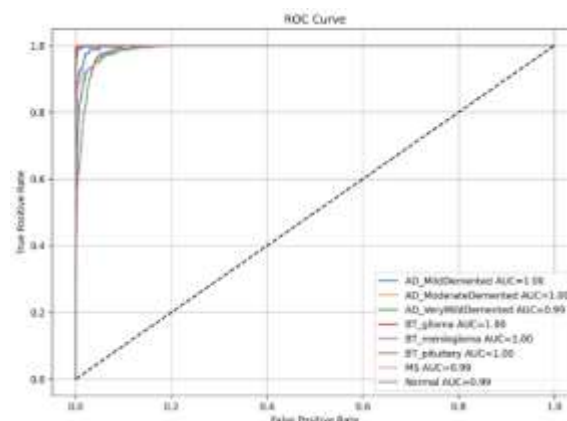


Figure 9: ROC Curve

The more detailed progression of training and validation accuracy, respectively, is shown in Figures 10 and 12;

the analogous progression of training and validation loss is demonstrated in Figures 11 and 13. The accuracy curves are improving steadily as the model is trained, and the loss curves show that the classification error is gradually decreasing during the optimization. The lack of a great discrepancy between training and validation indicators proves the advantage that the suggested framework has in preventing overfitting in the training process.

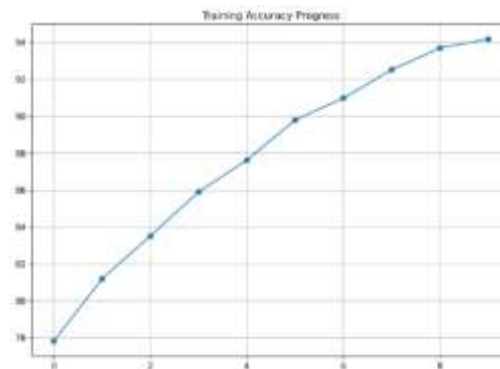


Figure 10: Training Accuracy

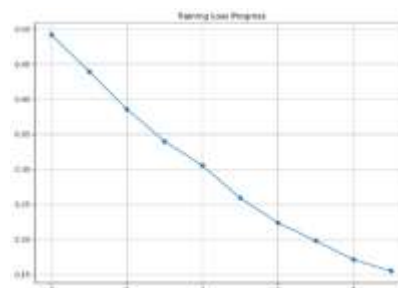


Figure 11: Training Loss

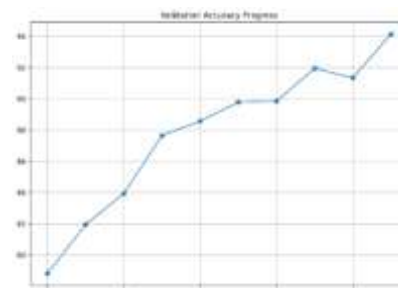


Figure 12: Validation Accuracy



Figure 13: Validation Loss

5. Conclusion

This paper introduced a Vision Transformer (ViT)-based model that was developed to recognize various neurological diseases automatically using medical-imaging datasets. The suggested model was successful in categorizing various neurological conditions, such as stages of Alzheimer's disease, brain tumours, multiple sclerosis, and healthy brain states. The framework managed to extract global contextual and local features in a neurological image, leading to better disease classification with the help of transfer learning on the pretrained vit_base patch16 224 architecture.

The experimental findings showed the proposed framework to be very predictive in a variety of neurological diseases. The training and validation curves were evidence of successful learning behaviour with low overfitting, and the confusion matrix and ROC-AUC analysis of the model attest to the high diagnostic power for neurological diseases. The framework also preserved effective performance when there is an imbalance in classes and complex neurological images.

Moreover, the intelligent medical image analysis framework proposed was efficient and scalable based on the computational resources of the Vision Transformer to analyse medical images. Comparative analysis and ablation study findings revealed that classical CNN-based methods were worse in comparison to the combination of pre-processing techniques and transfer learning in terms of classification performance. In general, the proposed framework is an efficient, extendable, and valid system of automated diagnosis of neurological diseases based on medical imaging data. The further directions of the work will involve the combination of Explainable AI, multimodal medical data analysis, and high-level deep learning methods that are likely to enhance interpretability, scalability, and clinical intelligent decision support capacity.

All the Declarations and Statements

Author Contributions Statement

All authors equally contributed to this work. All authors were involved in the research work, including methodology, experiments, analysis of the results, and preparation of the manuscript. All authors reviewed and approved the final manuscript.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Funding Declaration

This research did not receive any funding from any organization or funding agency. The work was carried out without external financial support.

Data Availability Statement

The dataset used in this study is publicly available on Kaggle and can be accessed at the following link: Multi-Class Neurological Disorder (MCND) Dataset. The dataset was used for research and experimental purposes in this study.

Ethical Declarations

This study used a publicly available dataset. No human or animal experiments were conducted in this work. Therefore, ethical approval was not required.

Acknowledgments

The authors would like to thank their institution and department for providing the necessary support and facilities to carry out this research work. The authors also thank the dataset providers for making the dataset publicly available for research purposes.

Declaration of Generative AI in Scholarly Writing

During the preparation of this manuscript, the authors used generative AI tools, including Grammarly and QuillBot, only for improving grammar, language, and readability. The authors carefully reviewed and edited the content generated or modified by these tools and take full responsibility for the accuracy, originality, and final content of the manuscript.

References

1. Sankari C, Jamuna V, Kavitha AR. Hierarchical multi-scale vision transformer model for accurate detection and classification of brain tumors in MRI-based medical imaging. *Scientific Reports*. 2025 Oct 31;15(1):38275.
2. Reddy C, Reddy PA, Janapati H, Assiri B, Shuaib M, Alam S, Sheneamer A. A fine-tuned vision transformer-based enhanced multi-class brain tumor classification using MRI scan imagery. *Frontiers in Oncology*. 2024 Jul 18;14:1400341.
3. Almansour AG, Alshomrani F, Almutairi AT, Alalwany E, Alshuhri MS, Alshaari H, Alfahaid A. DA-ViT: Deformable Attention Vision Transformer for Alzheimer's Disease Classification from MRI Scans. *Computer Modeling in Engineering & Sciences*. 2025;144(2):2395.
4. Tariq A, Iqbal MM, Iqbal MJ, Ahmad I. Transforming brain tumor detection empowering multi-class classification with vision Transformers and EfficientNetV2. *IEEE Access*. 2025 Mar 28.
5. Joy MA, Nasrin S, Siddiqua A, Farid DM. ViTAD: Leveraging modified vision transformer for Alzheimer's disease multi-stage classification from brain MRI scans. *Brain Research*. 2025 Jan 15;1847:149302.
6. Kaliappan M, Mariappan E, Manimaran V, Revathi B. Unified Vision Transformer for Multimodal Medical Image Classification. Available at SSRN 5281786. 2025.

7. Saha S, Akhter F, Tanny MR, Chowdhury NH. AGHC-neurotransformer: an attention-guided hybrid CNN-transformer model for robust classification of neurodegenerative disorders from MRI scans (Doctoral dissertation, Brac University).
8. Rahman Y, Hoque M, Chowdhury M, Al Mahmud M, Mahi MI. Cross-attention fusion vision transformer for explainable and efficient multi-class eye disease detection (Doctoral dissertation, BRAC University).
9. Moiz A, Kumar B. Supervised Deep Learning Framework for Multiclass Brain Tumor Detection Using MRI and Computer Vision Transformer. In 2026 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE) 2026 Jan 22 (pp. 1-6). IEEE.
10. Hussain T, Shouno H, Hussain A, Hussain D, Ismail M, Mir TH, Hsu FR, Alam T, Akhy SA. EFFResNet-ViT: A fusion-based convolutional and vision transformer model for explainable medical image classification. IEEE Access. 2025 Mar 26.
11. Sameera V, Meghana H, Ameer P, Parayangat M, Abbas M. Transformers for multi-modal image analysis in healthcare. Computers, Materials, & Continua. 2025;84(3):4259.
12. Chakra Borty J, Chakraborty GS, Noman IR, Batra S, Das J, Bishnu KK, Tarafder MT, Islam A. A novel diagnostic framework with an optimized ensemble of vision transformers and convolutional neural networks for enhanced Alzheimer's disease detection in medical imaging. Diagnostics. 2025 Mar 20;15(6):789.
13. Bhattacharjee S, Nath AR, Amin MR. Vision Transformer Based Approach for Multi-Class Eye Disease Classification. In 2025 28th International Conference on Computer and Information Technology (ICCIT) 2025, Dec 19 (pp. 688-692). IEEE.
14. Fasnamol AA, Sabu MK, Rehman KR. Explainable Weighted Cross-Attention Based Hybrid CNN-Transformer Network for Multimodal Brain Disease Diagnosis. In 2026 2nd International Conference on Advances in Intelligent Computing and Applications (AICAPS), 2026 Feb 11 (pp. 1-8). IEEE.
15. Draksharam M, Rao KV. A Transformative Approach for Multi-Class Glaucoma Detection Using Deep Learning Attention-Vision Transformer Model. Mathematical Modelling of Engineering Problems. 2025 Nov 1;12(11).
16. MR BR, Kavitha K. Multi-Class Neurodegenerative Disease Classifiers for Brain MRI Using Swin Transformer. In 2025 Innovations in Power and Advanced Computing Technologies (i-PACT) 2025 Sep 25 (pp. 1-8). IEEE.
17. Priyan SV, Nishok VS, Chinnadurai J, Nivethitha T. Next-generation sparse Transformer framework for precision detection and Dynamic stage classification of rare tumors in ultra-high-resolution medical imaging. Biomedical Signal Processing and Control. 2026 Jun 15;119:109832.
18. Chowdhury MF, Ferdous J, Korani W. Hybrid-CNNViT: A Deep Learning Framework for Multi-Class Brain Tumor Classification and Computer-Assisted Intervention Support. In Medical Imaging with Deep Learning 2026.
19. AlMohimeed A. Multi-stage framework using transformer models, feature fusion and ensemble learning for enhancing eye disease classification. Scientific Reports. 2025 Aug 19;15(1):30469.
20. Imran B, Riadi S, Basri H, Erniwati S, Muahidin Z, Efendi MM, Yani A. NeuroViT: A Neuro-Adaptive Vision Transformer with Predictive Coding for Interpretable Multi-Class Ocular Disease Classification. In 2025 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS) 2025 Dec 3 (pp. 752-760). IEEE.
21. Güven M. Detection of Alzheimer's and Parkinson's diseases using deep learning-based various transformer models. Engineering Proceedings. 2024 Oct 11;73(1):4.
22. Asuai CE. An Adaptive-Weighted Ensemble of CNNs, RNNs, and Vision Transformers for Multi-Modal Neuroimaging in Amyotrophic Lateral Sclerosis Diagnosis. Journal of Fuzzy Systems and Control. 2025 Dec 4;3(3):231-8.
23. Konatham NL, Thati J, Kadha V. Multi-Stage Feature Fusion using Vision Transformer with Residual CNN and Channel Attention for Enhanced Brain Tumor Detection. In 2025 5th International Conference on Artificial Intelligence and Signal Processing (AISP), 2025 Nov 22 (pp. 1-5). IEEE.
24. Panigrahi S, Adhikary DR, Pattanayak BK, Dash BB, De UC, Patra SS. Interpretable Multi-Class Brain MRI Tumor Classification with ViT-B16 Transformer. In 2025 IEEE 4th International Conference for Advancement in Technology (ICONAT), 2025 Sep 19 (pp. 1-5). IEEE.
25. Ahmed S. Dual-Stage Vision Transformer (DS-ViT) with Attention-Guided Patch Aggregation for Alzheimer's Disease Classification. In 2026 5th International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), 2026 Jan 29 (pp. 1-6). IEEE.
26. Ali A, Shahbaz H, Ganie SM, Alfuraydan MM. X-ViT-CNN: A Novel Network-Level Fusion of Transfer Learning and Customized Vision Transformer for Multi-Stage Alzheimer's Disease Prediction Using MRI Scans. Diagnostics. 2026 Mar 11;16(6):835.
27. Jeyaprabha B, Bist AS, Ch S, Verma RK, Pandey M. Dual Attention Enhanced Transformer Framework for the Multi-Class Alzheimer Diagnosis Using Brain MRI Images. In 2026 International Conference on Computing, Sciences and Communications (ICCS), 2026 Feb 12 (pp. 1-6). IEEE.
28. Aamir M, Rahman Z, Choudhry N, Bhutto JA, Abro WA, Zhu Z. From CNNs to Transformers: A Review of

- Evolving Deep Learning Architectures for Brain Tumor Classification. IEEE Access. 2025 Oct 27.
29. Alrahhah M, AlShabi M, Gadsden SA. MRI-Based Brain Tumor Classification Using Data-Efficient Vision Transformer (DeiT) for High-Precision Multi-Class Diagnosis.
30. Gómez-Guzmán MA, Jiménez-Beristain L, García-Guerrero EE, Aguirre-Castro OA, Esqueda-Elizondo JJ, Ramos-Acosta ER, Galindo-Aldana GM, Torres-Gonzalez C, Inzunza-Gonzalez E. Enhanced multi-class brain tumor classification in MRI using pre-trained CNNs and transformer architectures. Technologies. 2025 Aug 22;13(9):379.
31. Velu K, Jaisankar N. Design of a CNN-Swin transformer model for Alzheimer's disease prediction using MRI images. IEEE Access. 2025 Aug 25. <https://www.kaggle.com/datasets/alifatahi/multi-class-neurological-disorder-mcnd-dataset/data>