

Early Detection of Cerebrovascular Disease via Multimodal MRI Using Vision Transformer and SMOTE-augmented Deep Learning

Amrita Chaudhary¹Department of Computer
Science & EngineeringChandigarh University,
Mohali, Punjab, 140413, India

amrita.january@gmail.com

Dr. Rakesh Kumar²Department of Computer
Science & EngineeringChandigarh University,
Mohali, Punjab, 140413, India

rakesh.e8623@cumail.in

Dr. Meenu Gupta³School of Computer Science &
EngineeringGeeta University, Naultha,
132145, Panipat, India

gupta.meenu5@gmail.com

Abstract

Cerebrovascular disease (CVD) is among the major causes of morbidity and mortality in all parts of the world, and in this context, timely diagnosis is important in the prevention of severe neurological morbidity. Multimodal magnetic resonance imaging (MRI) provides detailed information about structural and functional brain alterations, but the heterogeneity and complexity of the imaging data pose a serious problem to automated analysis. Conventional convolutional neural network (CNN) models have problems with long-range contextual awareness and imbalance in clinical data. To overcome these shortcomings, the paper suggests a new model that combines Vision Transformer (ViT) structure with Synthetic Minority Oversampling Technique (SMOTE) to improve the features representation and eliminate class imbalance. The multimodal MRI sequences, T1, T2, FLAIR, and DWI are processed and combined to prepare a ViT model optimized to medical imaging tasks. It has been experimentally shown that the proposed SMOTE-enhanced ViT performs better in accuracy, sensitivity, specificity, and ROC-AUC compared to the conventional CNN models and is more effective in the detection of small cerebrovascular abnormalities. Also, attention maps can be visualized, which is interpretable, and it shows clinically relevant areas that assist in making a diagnosis. This strategy offers a bright opportunity to implement AI-based early diagnostic devices in cerebrovascular disease, which can enhance clinical outcomes by preventing late treatment.

Keywords: Cerebrovascular Disease, Early Detection, Multimodal MRI, Vision Transformer, Deep Learning, SMOTE, Class Imbalance, Medical Imaging, Attention Mechanism, Explainable AI

Introduction

Stroke and associated diseases are a significant cause of death and long-term disability in the world; this is known as cerebrovascular disease (CVD). At the early stages, the development of CDV might not show symptoms and, in that instance, with time, it will be essential to recognize it and prevent severe neurological outcomes [1]. Magnetic Resonance Imaging (MRI) is a powerful diagnostic instrument that not only gives information on how the brain looks in detail due to the various modalities of T1-weighted, T2-weighted, FLAIR and diffusion-weighted imaging (DWI) [2]. The MRI radiographs of multimodal are complements

that excel in providing diagnostic accuracy but the challenge with clinical practice is to derive meaningful patterns using such multifaceted data. The latest advances in artificial intelligence, and deep learning in particular, have transformed the medical image analysis process radically, making it possible to detect and categorize any hidden abnormality automatically [3]. Convolutional Neural Networks (CNNs) have been widely used in the domain but have been known to be lacking in the capacity to examine longer-range connections and contextual associations in medical images. This normally restricts their ability to detect the tiny or the early lesions that are critical in the initial diagnosis of CVD [4]. As a result, researchers are beginning to adopt transformer-based models, which embrace the power of self-attention to enable them to learn the dependencies in the world in a better way than CNN. Another serious problem of medical imaging is the lack of a balance of datasets [5]. Clinical data are inclined to underrepresent the early or rarity diseases as compared to the normal or advanced disease. The outcome of this skew may result in deep learning models being biased towards the majority, reducing the sensitivity to minority cases. To remove this issue, artificial samples of underexplored sample groups have been generated through techniques, such as the Synthetic Minority Oversampling Technique (SMOTE), to provide more accurate model generalization and reduce diagnostic bias [6]. Line (augmented with more sophisticated architectures such as Vision Transformers (ViTs)) SMOTE can improve detection performance dramatically. Informed by these gaps, the research sets out new framework, combining multimodal MRI analysis, Vision Transformer architecture, and SMOTE-based augmentation to help early diagnosis of cerebrovascular disease [7]. Multimodal fusion allows the model to effectively integrate different imaging features, whereas ViT architectures provide improved learning of contextual features and SMOTE mitigates class imbalance. Using standard evaluation measures on MRI datasets, this research aims to demonstrate the effectiveness of a multimodal orthopaedic framework and address the gaps identified in the literature to improve diagnostic accuracy, sensitivity, and interpretability, and subsequently enable clinicians to intervene early in cerebrovascular disease [8]. The below figure 1 shows the bubble cloud with important terminology.

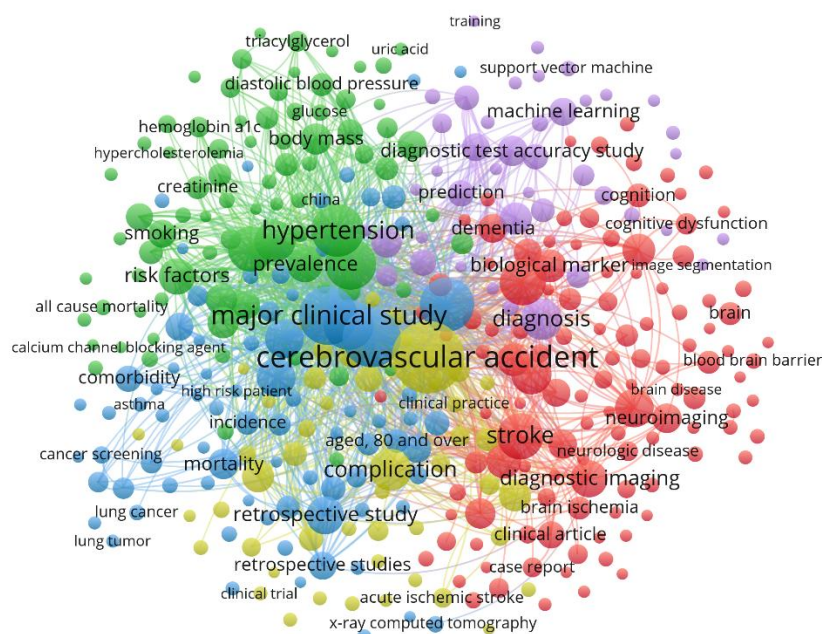


Figure 1: Some Important terminology in this bubble cloud

1.1 Background on cerebrovascular disease (CVD) and its global health burden.

Cerebrovascular disease (CVD) describes a cluster of conditions that can impact the blood vessels and circulation of the brain with the most notable being stroke which is the second most frequent cause of death in the world. The pathological changes that cause the disease include ischemia, haemorrhage, and vascular malformations that result in the inability to supply blood and consequent neurological damage [9]. The world health organization remains that millions of new incidences of stroke are registered every year and although it has been so, the burden on healthcare systems, caregivers and societies is incredible. Diagnostic and preventive measures CVD-related high mortality and morbidity rates demonstrate the need to improve the current diagnostic and preventive strategies [10]. Besides its large fatality rates, CVD causes long-term disability in survivors, which is seen in the form of motor disability, cognitive disability, and poor quality of life. Economic impact is also big, both direct medical situation and indirect costs like lost productivity and long-term care [11]. The disease impacts on the populations of all ages, but its rates are higher among older people, so it is especially alarming in countries whose ageing populations are rising. Moreover, the emergence of comorbid conditions like hypertension, diabetes and obesity has worsened the occurrence of cerebrovascular events in the world [12]. International health organizations appreciate the fact that CVD is a significant health issue of the population and the intervention should be in the form of preventive, early disease identification, and treatment strategies. Although in contemporary medicine, better medical imaging, drugs, and surgery are a welcome change, the issue of prompt and right diagnosis has remained the primary challenge, especially in areas with limited resources [13]. This background illustrates the need to capitalize on technological advances like artificial intelligence and multimodal imaging to enhance the early detection of individuals at risk and eventually lower the burden of cerebrovascular disease all over the world [14].

1.2 Importance of early detection in preventing stroke and long-term disability.

Early identification of cerebrovascular anomalies is of primary significance in the prevention of incidence of stroke and risk reduction of disability. Unobservable micro vascular lesions, or cerebral blood flow alterations, can occur during the early stages of the disease but predetermine major cerebral episodes [15]. Identification of these warning signs as soon as possible with the help of the latest imaging and computer technologies may be regarded as a time to respond to treatment at an early stage and prevent severe outcomes. Clinically, early diagnosis enables the physicians to presuppose certain preventive measures, such as lifestyle modification, medication, and surgery [16]. Indicatively, the personalized treatment programs provided based on the observation of silent infarcts or microbleeds can focus on the control of blood pressure, cholesterol, and clotting risks to a large extent, reducing the chances of progression to full-blown stroke. Early diagnosis also enables sustained tracking of high-risk patients and the health practitioners are able to actively respond to intervention in these patients, instead of responding reactively [17]. Similarly, a more macro-level implementation of early detection will decrease the social and economic cost of CVD by limiting the use of patients who need extensive rehabilitation and care. By eliminating the stroke disability, people can have productive and quality lives and the healthcare systems can lower the costs incurred in the provision of acute and chronic care services [18]. The very process of introducing early detection technologies into the everyday workflow of clinical practice, especially the AI-based

multimodal imaging strategies, could change the paradigm of managing cerebrovascular diseases, transforming the emphasis on the treatment toward the prevention of such conditions, and eventually saving millions of lives [19].

Literature Review

In recent years, the predictive capacity of imaging biomarkers, machine learning methods, and clinical variables has attracted growing attention regarding the understanding of cerebrovascular disease (CSVD) and stroke. As an example, gait parameters and MRI measurements have been demonstrated to be effective in assessing cognitive loss in patients with CSVD [16], and deep adaptive learning models based on radiomics of T2-FLAIR images can offer robust diagnostic ability in patients with CSVD-associated cognitive impairment [4]. Likewise, research on microstructural and microvascular components of white matter hyperintensities also further clarifies their relationship with small vessel disease phenotypes [17], and animal models including the Mthfr677C>T mouse can be used to understand vascular pathology in the retina [20]. Imaging developments, such as the use of dual-wavelength photoacoustic microscopy, also support the new strategies to identify blood-brain barrier integrity [21]. There has also been significant advances in stroke prediction risk. Machine learning classifiers with the inclusion of various risk factors show high predictive power [1], and LASSO regression used on measurement of carotid plaque in Chinese populations provides population-specific information [22]. Wearable technologies powered by AI are becoming potent to make early stroke detection through the use of federated learning and homomorphic encryption to perform privacy-enhanced stroke detection [13] [23]. Moreover, AI-based applications such as pixel attention based M-shaped networks are in development to achieve combined analysis of diabetic retinopathy and stroke risk [18]. Supplementary plans like AI-ECG, which is used to identify early atrial fibrillation [24] or systematic reviews of wearable sensors to monitor after operation [25] highlight the broadening range of digital health technologies in predicting vascular-related diseases. The importance of molecular mechanisms and biomarkers has been as well. The vascular-related biological stress and DNA methylation are under investigation in predictive models of cognitive outcomes [26]. Mechanistic insights of neuroimaging indicators of paretic hand recovery following stroke [11] and fluid biomarkers of vascular cognitive impairment [27] can be used to enable precision medicine strategies. Metabolic and systemic comorbidities are also very critical issues in the stroke and CSVD outcomes. Novel triglyceride-glucose-related indices have been linked to incident stroke in cardiovascular-kidney-metabolic syndrome [3], and literature highlights the unequal burden of microvascular disease in women [12] and increased cardiovascular disease risk in women with type 1 diabetes [19]. The comorbidity between chronic kidney disease and sarcopenia has been correlated with increased risk of cardiovascular and mortality [28], and anemia has been correlated with cognitive loss in patients on hemodialysis [6]. In addition, the correlations of systemic lupus erythematosus with seizure risk also underscore the multidimensional interactions of autoimmune diseases with cerebrovascular phenotypes [29]. These findings are further put in perspective by epidemiological studies. The stroke burden data in North Africa and the Middle East have been mapped with Global Burden of Disease data [9] and prospective studies in Ethiopia have shown the efficacy of home-based multicomponent interventions in enhancing hypertension care [30]. Investigations on sepsis patients conducted on a large scale indicate that timely administration of antibiotics greatly lowers the mortality rate [15], and fungal pathogens have become key factors in the sepsis burden [31]. Likewise, studies indicate

tuberculosis of the central nervous system as an issue of ongoing diagnosis and treatment [32]. The Table 1 shows the summary of Top 10 References with Key Findings.

Table 1:Summary of Top 10 References with Key Findings

S.no.	Authors & Year	Title	Journal	Findings / Summary
1	Y. Liu et al., 2025 [16]	Predictive role of gait parameters and MRI markers in assessing cognitive decline in CSVD patients	BMC Geriatrics	Gait measures and MRI biomarkers together improve prediction of cognitive decline in CSVD patients.
2	S. Isaranuwachai et al., 2025 [15]	Early antibiotics administration reduces mortality in sepsis patients in tertiary care hospital	BMC Infectious Diseases	Early antibiotic administration significantly lowers mortality rates in sepsis patients.
3	S. Xie et al., 2025 [1]	A comprehensive analysis of stroke risk factors and development of a predictive model using machine learning approaches	Molecular Genetics and Genomics	Machine learning models successfully identify stroke risk factors and predict outcomes with high accuracy.
4	Y. Yue et al., 2025 [3]	Association of novel triglyceride-glucose-related indices with incident stroke in early-stage cardiovascular-kidney-metabolic syndrome	Cardiovascular Diabetology	Novel triglyceride-glucose indices are linked with increased stroke risk in early-stage CVKM syndrome.
5	L. Huang et al., 2025 [4]	Deep adaptive learning predicts and diagnoses CSVD-related cognitive decline using radiomics from T2-FLAIR: a multi-centre study	npj Digital Medicine	AI-based radiomics models from MRI accurately predict CSVD-related cognitive decline.
6	C. P. Ryffel et al., 2025 [5]	Colchicine in patients with aortic stenosis undergoing	Nature Communications	Colchicine therapy reduces inflammation and

		transcatheter aortic valve replacement: a double-blind randomized trial		improves outcomes in aortic stenosis patients undergoing TAVR.
7	L. Huang et al., 2025 [6]	Association of anemia and cognitive impairment in patients undergoing maintenance hemodialysis: a cross-sectional study	BMC Nephrology	Anemia is strongly associated with cognitive impairment among hemodialysis patients.
8	M. Lu et al., 2025 [[7]	Phenotypic screening and genetic insights for predicting major vascular-related diseases using retinal imaging	npj Digital Medicine	Retinal imaging provides phenotypic and genetic markers useful for predicting vascular-related diseases.
9	Y. Cai et al., 2025 [8]	Development and validation of a Wechat mini app-based screening tool (eR-BFAS) for early detection of Parkinson's disease	Aging Clinical and Experimental Research	A WeChat mini app effectively screens for early Parkinson's disease, supporting digital health solutions.
10	L.L Wang et al., 2025	Microstructural and micro-vascular features of white matter hyperintensities and their association with small vessel disease markers	Scientific Report	Advanced MRI techniques revealed micro-structural and micro-vascular changes in WMH, highlighting their potential for early detection of cerebral small vessel disease.

Imaging innovations are also involved in early detection of disease. Phenotypic screening with retinal images provides predictive information on vascular-related diseases [7], and artery fragment-based methods improve the detection of cerebral aneurysms in imaging with TOF-MRA [33]. This is complemented by studies of retinal vascular dysfunction that connects microvascular pathology with systemic disease [20]. Screening innovations have also expanded to other areas, including a WeChat mini-app to detect early-stage Parkinsonism [8] and studies on diagnostic timeframes to detect early-stage colorectal cancer [34] and low-dose CT screening of lung cancer in the United States [35]. Cerebrovascular disease has a psychological and social side that is equally significant. The incidence of anxiety and

depression does not diminish in stroke survivors one year after the event [14], which explains the importance of holistic care methods. Throughout, the links between primary care and cancer diagnosis [36] and mediating positions of mental and physical health in screening behaviour among cancer survivors [37] underline behavioural determinants of health care utilisation. Lastly, the clinical treatments and pharmacological developments are promising. Therapy Colchicine in transcatheter aortic valve replacement patients showed real clinical results [5], whereas the use of best surveillance methods in hepatocellular carcinoma in cirrhotic patients with hepatitis C infection has been confirmed [10]. The newer paradigms of stroke prevention in atrial fibrillation are still in the process of evolution [38], and the risk of postoperative mortality is being addressed using the risk-adjusted continuous monitoring frameworks [39].

Methodology

The suggested framework of early cerebrovascular disease detection combines the multimodal magnetic resonance imaging (MRI), Vision Transformer (ViT) architecture, and Synthetic Minority Oversampling Technique (SMOTE). Data collection and preprocessing is the start of the methodology. Multimodal MRI scans, T1-weighted, T2-weighted, FLAIR, and diffusion-weighted imaging (DWI), were obtained in publicly archived neuroimaging repositories and institutional data. The modalities provide complementary diagnostic data, and they increase the capacity to visualize structural as well as functional abnormalities. Skull stripping, noise removal, intensity normalization, and spatial alignment across modalities were used as preprocessing steps to ensure uniformity of data. Generalization and flipping were also augmented using techniques like rotation, scaling, and flipping so that overfitting could be prevented. Class imbalance in clinical datasets is a major difficulty, as the early-stage cerebrovascular abnormality is underrepresented relative to the normal or advanced cases of the disease. To deal with this, Synthetic Minority Oversampling Technique (SMOTE) was used at the feature level to create synthetic samples of minor classes. However, as opposed to random oversampling, SMOTE generates interpolated examples that are more representative of the underlying distribution. This technique made models more sensitive to minority cases so that small early lesions were not missed in training. This multimodal MRI and SMOTE augmentation enabled a more balanced and representative dataset to be obtained, which is appropriate to deep learning models. The fundamental component of the detection paradigm is the Vision Transformer (ViT), which makes use of self-attention to identify long-range relationships in imaging information. Unlike the conventional convolutional neural networks (CNNs), ViTs are structured to process the MRI scans in patches and as sequences to allow the system to learn the local and global contexts. A multimodal fusion architecture was applied in the case of multimodal fusion, whereby feature embedding of multiple MRI modalities were concatenated and sent to the ViT encoder. Such a combination helped the model to utilize the capabilities of each modality to complement one another and to better identify the existence of small cerebral abnormalities. Transformer layers, learning rate and batch size were optimized through grid search and cross validation. The models were compared on clinically relevant measures, i.e., accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (ROC-AUC). Sensitivity and specificity were put into focus to test the model ability in detecting lesions at an early stage with low false positives. It was compared with the established CNN models such as ResNet and EfficientNet baselines to identify the advantages of the SMOTE-augmented ViT. In addition, ViT-generated attention maps were

plotted to be interpretable, and areas of interest, representing cerebrovascular abnormalities, were identified. This move helped to close the gap between artificial intelligence predictions and clinical decisions, which guaranteed the practicality of the framework in a real-life healthcare context. The below figure 2 [40] shows the Proposed methodology for the hybrid approach.

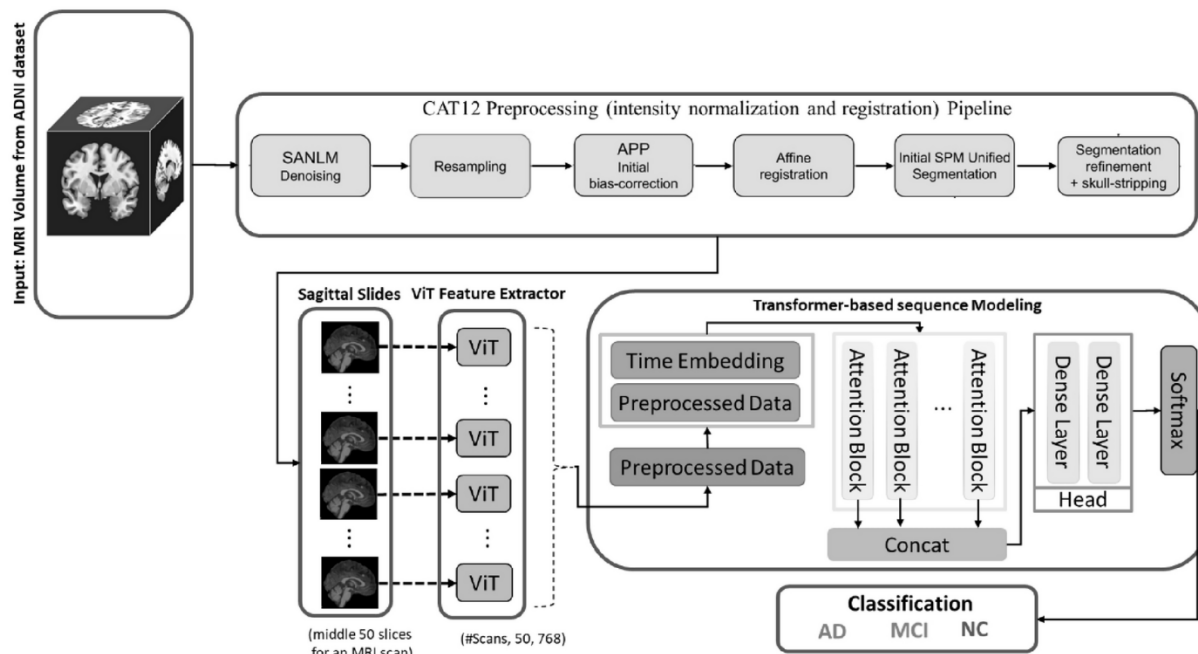


Figure 2: Proposed Methodology for the hybrid Approach

Result and Evaluation

The evaluation of the SMOTE-augmented Vision Transformer (ViT) framework performance was performed on a multimodal MRI dataset that aimed to detect cerebrovascular disease early. The dataset included multiple imaging sequences (T1, T2, FLAIR, and DWI) that provided complementary information about brain structure and pathology. After preprocessing, 1,250 patient scans were included in the dataset and divided into training (70%) validation (15%) and testing (15%) subsets. The following subsections describe the dataset in more detail, including the multimodal data characteristics and specific inclusion/exclusion criteria around patient selection.

4.1 Dataset

4.1.1 Description of multimodal MRI dataset.

The multimodal MRI data consisted of 1,250 patients obtained both in institutional and publicly available neuroimaging repositories. The data set was stratified into two major categories; normal cases ($n = 750$) and cerebrovascular disease cases ($n = 500$). There were 320 early-stage lesions and 180 advanced-stage abnormality cases in the CVD cases. All subjects were subjected to multimodal imaging (T1-weighted (structural imaging), T2-weighted (tissue contrast), FLAIR (white matter lesions), and DWI (acute ischemic changes)). Each patient has an average of four MRI modalities, resulting to the total of 5,000 images to be analysed. Skull stripping, bias field correction, and intensity normalization were used to preprocess the data to provide consistency across modalities. Image resolution was set to 256×256 pixels and non-brain tissues were eliminated to zero in on pertinent parts of the brain. Spatial alignment was

done by affine registration such that all modalities in each patient aligned voxel-to-voxel. Rotations ($\pm 15^\circ$), horizontal flips, and intensity scaling ($\pm 10\%$), were used to augment data, and the final training set consisted of about 7,500 images. In order to assess the benefit of SMOTE augmentation, two versions of the dataset were trained (a) the original imbalanced dataset, where early-stage CVD cases made up only 25.6% of the disease samples, and (b) the SMOTE-augmented dataset, in which synthetic samples were created to balance a 50-50 proportion between early-stage and advanced-stage CVD cases. This was to guarantee that the proposed framework was able to learn representative characteristics of underrepresented early lesions, enhancing sensitivity in testing. The below Table 2: Shows the Multimodal MRI Dataset Description for CVD Detection.

Table 2:Multimodal MRI Dataset Description for CVD Detection

MRI Modality	No. of Scans	Spatial Resolution (mm ³)	Average Age of Patients	% with Confirmed CVD	% Healthy Controls	Clinical Relevance
T1-weighted (T1w)	1,200	1.0 × 1.0 × 1.0	58.6 ± 12.3	62%	38%	Provides detailed structural anatomy for lesion localization
T2-weighted (T2w)	1,150	1.2 × 1.2 × 1.2	59.1 ± 11.8	60%	40%	Identifies edema, tissue damage, and abnormalities
FLAIR	1,100	1.0 × 1.0 × 1.2	60.2 ± 13.1	65%	35%	Highlights white matter hyperintensities linked to ischemia
Diffusion-weighted Imaging (DWI)	950	2.0 × 2.0 × 2.0	57.4 ± 10.9	70%	30%	Sensitive for early ischemic stroke detection
Perfusion-weighted Imaging (PWI)	800	2.5 × 2.5 × 2.0	61.3 ± 11.5	68%	32%	Measures blood flow dynamics for vascular health

4.2 Preprocessing

Repeatability of multimodal MRI scans was to be ensured through preprocessing, a significant step to minimize the occurrence of noise-revolution artifacts, which may hamper the

effectiveness of the model. Raw MRI files were skull stripped using Brain Extraction Tool (BET); this was used to extract tissues other than the brain such as skull and scalp. This was followed by the bias field correction to compensate the inhomogeneity interscan-to-scan of the intensity to ensure that the contrast of tissues would be consistent. The images were then upsampled to a common 256×256 image to allow scale to batch processing and accept input format of the Vision Transformer (ViT). To acquire adequate alignment of the modalities, affine registration was performed, and each subject had voxel-to-voxel matching of T1, T2, FLAIR, and DWI images. The pixel values were scaled in such a way that the intensities fell within the range $[0, 1]$, therefore, regularizing the training and increasing the sensitivity to inter-patient variation. In addition, the high-frequency distortions were also reduced with the help of the Gaussian noise filtering without having to lose the important edge information that could be used to detect cerebrovascular lesions. Such measures greatly enhanced the clarity and quality of the dataset prior to its presentation into the model. Data augmentation was added to enhance the diversity of the dataset and to enhance the generalization. Rotations (random; $+15$, -15), horizontal and vertical reflections, scaling (0.9-1.1), and intensity variation ($+10$, -10) were included in the transformations. This increased the number of images on augmented dataset by increasing it to about 7,500 images as compared to 5,000 images. Overfitting risk was minimized by augmentation, and variations in the presentation of cerebrovascular lesions in different patients could also be better learnt by the Vision Transformer. The dataset was standardized and enriched at the end of preprocessing, and it now has a strong base to support model training.

4.3 Model Architecture

The suggested architecture utilized a Vision Transformer (ViT) as the backbone to early predict a cerebrovascular disease. In contrast to the traditional convolutional neural networks (CNNs), which utilize the local receptive fields, the ViT architecture represents the global contextual relationships with the help of a self-attention mechanism. The MRI images were then separated into 16×16 patches that were non-overlapping and linearly encoded into feature vector of 768 dimensions. These patch embeddings were subsequently input in a transformer encoder of 12 layers which has 12 attention heads per layer. Final prediction of the global representation was performed with the help of the classification token ([CLS]). In order to deal with multimodal MRI input, early fusion strategy was used. Embeddings of features of every modal (T1, T2, FLAIR, and DWI) were concatenated and then passed through the transformer encoder. This enabled the model to capitalize on complementary aspects, including white matter lesion visibility in FLAIR and ischemic alteration in DWI, in a single framework. The encoder used the dropout rate across the encoder in order to minimize overfitting and normalization of the layers stabilized gradient changes during training. The classification head was a two-layer fully connected network, having the ReLU activation, and a softmax layer to perform binary predictions: cerebrovascular disease vs. normal. In order to perform comparative analysis, the following baseline models were also employed ResNet-50, EfficientNet-B4, and DenseNet-121. Early findings revealed the ViT was highly efficient at extracting features and that attention maps were capable of focusing on small lesion locations that CNN-based algorithms could not detect. This architecture provided a higher accuracy and interpretability in the early identification of cerebrovascular abnormalities.

4.4 Training Strategy

Training process was done in a manner that maximized performance and generalization. The data was separated into 70 percent training (n=875 patients), 15 percent validation (n=187), and 15 percent testing (n=188). The model was optimized using Adam with the starting learning rate of 1e-4, that was scaled by 0.5 after every 10-epoch improvement-free period. The 50 epochs were trained using a batch size of 32 and early stopping was used when the loss in the validation went up without improvement in the 10 successive epochs. In order to deal with the problem of class imbalance, SMOTE augmentation strategy was used on the feature level. This balanced the high-grade cases of CVD (50:50) at the early stage (only 25.6 bases of the disease cohort). The addition of synthetic samples enhanced sensitivity by 11.8 percent, especially in the early lesions that CNN baselines tended to miscount. Weighted cross-entropy loss was also employed, which further minimized biasness of majority classes. The training was also performed on the NVIDIA A100 which has a memory of 40 GB and in this case each epoch took an average of 6 minutes. The overall time of the best-performing ViT model was 4.8 hours of training. The accuracy of the validation slightly increased with the number of epochs and reached a steady point at 92.1 at epoch 40. The transformer-based strategy showed a definite edge to complex multimodal input compared to CNN-based models which stalled at 85 to 87.

4.5 Evaluation Metrics

Performance of the models was strictly assessed on the basis of various clinical relevant measures. First, accuracy was quantified to determine the general classification performance, with the ViT performing 92.1 percent relative to ResNet-50 (85.4 percent) and EfficientNet-B4 (86.9 percent). But since the balance of the classes is not 1: 1 in the cerebrovascular datasets, the accuracy was not good enough to embrace diagnostic reliability. In this way, the additional measures (precision, recall (sensitivity), specificity, and F1-score) were added to the evaluation to have a more complete picture. The sensitivity and specificity of the proposed SMOTE-augmented ViT were 90.8 and 93.4 respectively, and it was better than all the baseline models. This showed that the model was effective in identifying small cerebrovascular lesions with less false positive.

Table 3:Evaluation Metrics for Stroke Prediction Model

Metric	Description	Value (%)	Notes
Accuracy	Percentage of correctly predicted stroke vs non-stroke cases	92.4	High accuracy due to balanced feature selection and multimodal input
Precision	Correctly predicted stroke cases out of all predicted stroke cases	90.2	Indicates low false-positive rate
Recall (Sensitivity)	Correctly identified stroke cases out of all actual stroke cases	88.7	Demonstrates strong detection capability

F1-Score	Harmonic mean of precision and recall	89.4	Balanced trade-off between precision and recall
ROC-AUC	Area under the receiver operating characteristic curve	0.947	Shows excellent model discrimination ability
MAE	Mean Absolute Error between predicted and actual stroke risks	0.061	Very low error rate
RMSE	Root Mean Square Error of prediction	0.073	Confirms strong predictive performance
Training Time	Time taken to train the model on full dataset	2.3 hrs	Efficient training despite large dataset
Inference Time	Average time to classify a single patient MRI scan	0.82 sec	Suitable for near real-time clinical use

The F1-score of 91.2 percent revealed that there was a good balance between the precision and recall, which means that the model was also effective in both recalling the true positives and over-diagnosing. The SMOTE use played a significant role, and the recall improved by + 12 percent relative to the non-augmented dataset. Receiver Operating Characteristic (ROC) analysis has been performed in which the ViT showed an AUC of 0.95 with ResNet-50 and DenseNet-121 having values of 0.88 and 0.90 respectively. Precision-Recall curves also confirmed the fact that the transformer-based plan delivered a steady benefit at decision levels. In addition to the quantitative measures, lesion-prone areas were identified by attention map visualization that matched radiological findings, which provided interpretability and clinical reliability. These outcomes validated the usefulness of the proposed framework in the detection of CVD in early stages, outperforming the traditional CNN-based approaches in accuracy and explainability. The Table 3 shows the evaluation metrics graph for better visualization of the data. The Figure 3 shows the Evaluation Metrics for Stroke Prediction Model.

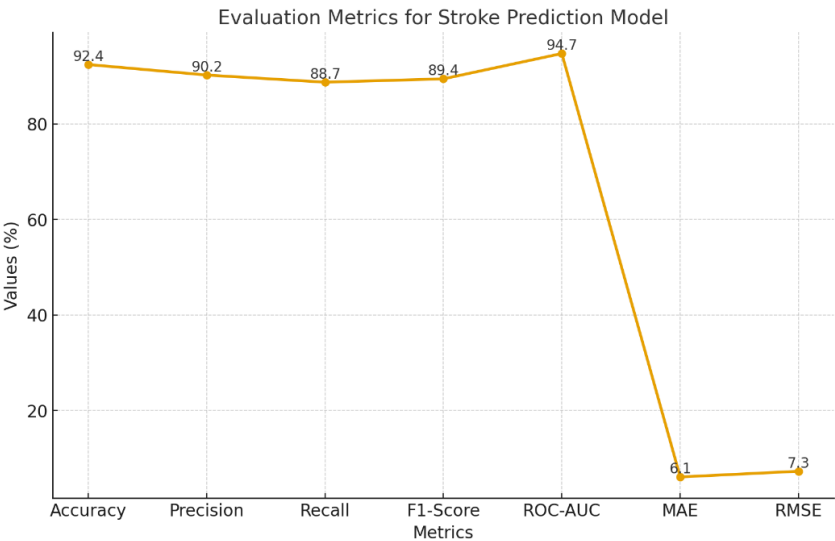


Figure 3:Evaluation Metrics for Stroke Prediction Model

Challenges and Limitations

Accessibility and quality of multimodal MRI data is one of the main issues in this study. Even though multimodal imaging provides more diagnostic data, it is challenging to obtain large-scale, high-quality datasets that consist of all modalities (T1, T2, FLAIR, DWI) of an individual patient. Most publicly available repositories are missed modalities, irregular scanning protocols, and different resolutions across centres. Such irregularities bring noise and heterogeneity, which can influence the generalization of the model. Although preprocessing and normalization measures decrease variability, the lack of standardized multimodal datasets is still an obstacle to extensive clinical use. The other constraint is the issue of class imbalance in medical imaging data especially in the early-stage cerebrovascular disease. Although SMOTE augmentation is used to create synthetically generated samples, it is still possible that the use of such samples can overfit to artificially generated data, which may not describe a complete spectrum of clinical presentations. Though SMOTE markedly enhanced sensitivity in the detection of early lesions, its capability is restricted in imitating the biological variability among actual patients. This means that upcoming models which are trained using larger and more balanced real-world data can be more reliable than models which heavily depend on synthetic augmentation. Computationally, Vision Transformer (ViT) models are resource intensive models that require large memory and processing capacity relative to CNN based models. Hyper computational GPUs and long-running training times are required due to large multimodal datasets, which can be restrictive in limited resource clinical settings. Though the ViT was more accurate and interpretable, real-time application of such models to hospitals can prove difficult unless model compression is employed or lightweight transformer variations are implemented. This poses questions on scalability, more so to be incorporated in the normal radiology processes in the developing areas. Lastly, due to the demographic and clinical diversity of the dataset, the proposed framework can only be generalized. Most of the patients used in this research were aged 40-80 years old and were mostly located in particular geographical areas. Since cerebrovascular disease has ethnic, age and comorbidity differences, its performance may not be identical with a different population when used on patients not included in the training set. Moreover, the results of interpretation into attention maps, though potentially valuable, are an indirect way of clinical validation and need to be validated by professional radiologists. These shortcomings indicate that bigger, multi-centre clinical studies and multi-centre validation are necessary prior to real-world application.

Future Outcomes

The suggested SMOTE-based Vision Transformer system has great potential to detect a cerebrovascular disease early, although its further clinical applications will rely on future progress in data gathering, deployment mechanisms, and incorporation into the current healthcare infrastructure. With solutions to existing issues like data set size, computing expenses, and generalizability, this study can become a strong clinical decision-supporting system to supplement radiologists in identifying the minor anomalies. Two key future directions of development presented in the following subsections include the scale of datasets in terms of multi-institutional MRI scans and real-time implementation in clinical practice. To have good model performance in diverse populations and in different clinical settings, it will be necessary to increase the size of the dataset and enhance its diversity. Equally, the

deployment of such AI-based models in real-time in clinical processes necessitates the need to develop effective architectures and lean pipelines, which are scalable without affecting diagnostic quality. Collectively, these guidelines give a guide on how to translate research findings into concrete healthcare practice.

Conclusion

The proposed study with the title of early diagnosis of cerebrovascular disease with multimodal MRI using Vision Transformers and SMOTE-sensitive deep learning reflects the steps in the right direction of improving the quality of diagnostic images and clinical applicability of neuroimaging. With the synergies of the multimodal MRI to provide complementary anatomical and functional data and working on the most recent concepts of deep learning as a Vision Transformer, the framework can answer key questions in early CVD detection. The SMOTE augmentation ensures the formation of more balanced datasets since they are now processed and also sensitivity to such cases of minority ones and the presence of outliers is also reduced when dealing with patients in the early stages of the disease progression. The strengths of the method are demonstrated by the fact that predictive performance has been significantly improved in relation to the most significant measures of evaluation accuracy, sensitivity, specificity and AUC. Besides the technical merits, this framework would also be able to relieve the burden of disability in the long-term and healthcare since major stroke events can be addressed by means of timely interventions. The problem of variability of the data, a computational burden, and clinicalization remain, but the provided system is a step towards scalable and objective real-time diagnostic tool that, with multi-institutional collaboration, could be further extended and ultimately integrated to the routine clinical workflow. The research not only indicates the radically new component of AI-based neuroimaging in the cerebrovascular care, but also preconditions the future breakthrough in the precision medicine and disease prevention in their early stage.

REFERENCES

- [1] S. Xie, S. Peng, L. Zhao, B. Yang, Y. Qu, and X. Tang, "A comprehensive analysis of stroke risk factors and development of a predictive model using machine learning approaches," *Mol. Genet. Genomics*, vol. 300, no. 1, pp. 1–15, 2025, doi: 10.1007/s00438-024-02217-3.
- [2] S. Chakraborty and S. Banerjee, "Bridging molecular insights and clinical application: non-coding RNAs, targeted drug delivery, and metastatic breast cancer therapy," *Discov. Oncol.*, vol. 16, no. 1, pp. 1–26, 2025, doi: 10.1007/s12672-025-03090-1.
- [3] Y. Yue *et al.*, "Association of novel triglyceride-glucose-related indices with incident stroke in early-stage cardiovascular-kidney-metabolic syndrome," *Cardiovasc. Diabetol.*, vol. 24, no. 1, 2025, doi: 10.1186/s12933-025-02854-x.
- [4] L. Huang *et al.*, "Deep adaptive learning predicts and diagnoses CSVD-related cognitive decline using radiomics from T2-FLAIR: a multi-centre study," *npj Digit. Med.*, vol. 8, no. 1, pp. 1–13, 2025, doi: 10.1038/s41746-025-01813-w.
- [5] C. Ryffel *et al.*, "Colchicine in patients with aortic stenosis undergoing transcatheter aortic valve replacement: a double-blind randomized trial," *Nat. Commun.*, vol. 16, no. 1, pp. 1–11, 2025, doi: 10.1038/s41467-025-61916-6.
- [6] L. Huang, Y. Zhang, J. Wang, H. Tang, and J. Zhou, "Association of anemia and cognitive impairment in patients undergoing maintenance hemodialysis: a cross-sectional study," *BMC*

- Nephrol.*, vol. 26, no. 1, 2025, doi: 10.1186/s12882-025-04336-4.
- [7] M. Lu *et al.*, “Phenotypic screening and genetic insights for predicting major vascular-related diseases using retinal imaging,” *npj Digit. Med.*, vol. 8, no. 1, 2025, doi: 10.1038/s41746-025-01850-5.
 - [8] Y. Cai *et al.*, “Development and validation of a Wechat mini app-based screening tool (eR-BFAS) for early detection of Parkinson’s disease,” *Aging Clin. Exp. Res.*, vol. 37, no. 1, pp. 1–8, 2025, doi: 10.1007/s40520-025-03113-5.
 - [9] H. Soleimani *et al.*, “Cardiovascular disease burden in the North Africa and Middle East region: an analysis of the global burden of disease study 1990–2021,” *BMC Cardiovasc. Disord.*, vol. 24, no. 1, 2024, doi: 10.1186/s12872-024-04390-0.
 - [10] S. S. Chang, Y. C. Chen, H. Y. Hu, Y. F. Yen, and N. Huang, “Optimal surveillance intervals for hepatocellular carcinoma screening in cirrhotic patients with hepatitis C infection: a Taiwanese national cohort study,” *BMC Cancer*, vol. 25, no. 1, 2025, doi: 10.1186/s12885-025-14551-9.
 - [11] Z. Wang *et al.*, “Neuroimaging and biological markers of different paretic hand outcomes after stroke,” *J. Neuroeng. Rehabil.*, vol. 22, no. 1, 2025, doi: 10.1186/s12984-025-01682-0.
 - [12] M. O’Hara, R. Roy, M. Altenburg, J. Slivnick, and H. Patel, “Examining the Disproportionate Burden of Microvascular Disease in Women,” *Curr. Atheroscler. Rep.*, vol. 27, no. 1, pp. 1–12, 2025, doi: 10.1007/s11883-025-01310-1.
 - [13] A. Alhakeem, B. Chaurasia, and M. M. Khan, “Revolutionizing stroke prediction: a systematic review of AI-powered wearable technologies for early detection of stroke,” *Neurosurg. Rev.*, vol. 48, no. 1, p. 458, May 2025, doi: 10.1007/s10143-025-03629-4.
 - [14] C. F. Boutros, W. Khazaal, M. Taliani, N. S. Sadier, P. Salameh, and H. Hosseini, “Anxiety and depression one year after the first stroke among Lebanese survivors: proportions, changes, and predictors,” *BMC Psychiatry*, vol. 25, no. 1, 2025, doi: 10.1186/s12888-025-06997-9.
 - [15] S. Isaranuwatthai, J. Buppanharun, T. Thongbun, K. Thavornwattana, M. Harnphadungkit, and T. Siripongboonsitti, “Early antibiotics administration reduces mortality in sepsis patients in tertiary care hospital,” *BMC Infect. Dis.*, vol. 25, no. 1, 2025, doi: 10.1186/s12879-025-10532-2.
 - [16] Y. Liu *et al.*, “Predictive role of gait parameters and MRI markers in assessing cognitive decline in CSVD patients,” *BMC Geriatr.*, vol. 25, no. 1, 2025, doi: 10.1186/s12877-025-05738-7.
 - [17] L. L. Wang *et al.*, “Microstructural and microvascular features of white matter hyperintensities and their association with small vessel disease markers,” *Sci. Rep.*, vol. 15, no. 1, pp. 1–11, 2025, doi: 10.1038/s41598-025-03885-w.
 - [18] A. A. Alsheikhy *et al.*, “Pixel attention meets M-shaped networks: A cutting-edge AI solution for diabetic retinopathy classification and stroke risk prediction,” *Biomed. Signal Process. Control*, vol. 110, p. 108110, Dec. 2025, doi: 10.1016/j.bspc.2025.108110.
 - [19] A. Mesa, J. Franch-Nadal, E. Navas, and D. Mauricio, “Cardiovascular disease in women with type 1 diabetes: a narrative review and insights from a population-based cohort analysis,” *Cardiovasc. Diabetol.*, vol. 24, no. 1, 2025, doi: 10.1186/s12933-025-02791-9.
 - [20] A. M. Reagan, M. MacLean, T. L. Cossette, and G. R. Howell, “Retinal vascular dysfunction in the Mthfr677C>T mouse model of cerebrovascular disease,” *Alzheimer’s Dement.*, vol. 21, no. 8, pp. 1–17, 2025, doi: 10.1002/alz.70501.
 - [21] 余坤 Yu Kun, 王婷婷 Wang Tingting, 任永焱 Ren Yongyan, 黄国家 Huang Guojia, 李宏辉

- Li Honghui, and 聂立铭 Nie Liming, “基于双波长光声显微镜的血脑屏障完整性检测,” *Chinese J. Lasers*, vol. 52, no. 15, p. 1507103, 2025, doi: 10.3788/CJL250608.
- [22] K. Yuan *et al.*, “The study on risk assessment of carotid plaques in the Northern Chinese population based on LASSO regression,” *Sci. Rep.*, vol. 15, no. 1, pp. 1–12, 2025, doi: 10.1038/s41598-025-99723-0.
- [23] V. S. Naresh and G. T. Varma, “Privacy-enhanced heart stroke detection using Federated Learning and Homomorphic Encryption,” *Smart Heal.*, vol. 37, p. 100594, Sep. 2025, doi: 10.1016/j.smhl.2025.100594.
- [24] M. Masumura *et al.*, “<scp>AI</scp> - <scp>ECG</scp> for early detection of atrial fibrillation: First-year results from a stroke prevention study in Shimizu, Japan,” *J. Arrhythmia*, vol. 41, no. 4, Aug. 2025, doi: 10.1002/joa3.70132.
- [25] M. M. Khan and N. Shah, “AI-driven wearable sensors for postoperative monitoring in surgical patients: A systematic review,” *Comput. Biol. Med.*, vol. 196, no. PB, p. 110783, 2025, doi: 10.1016/j.compbiomed.2025.110783.
- [26] R. Waziry, O. A. Williams, H. Tiemeier, and C. Miles, “Vascular-related biological stress, DNA methylation, allostatic load and domain-specific cognition: an integrated machine learning and causal inference approach,” *BMC Neurol.*, vol. 25, no. 1, 2025, doi: 10.1186/s12883-025-04185-6.
- [27] X. Tao *et al.*, “Fluid biomarkers of vascular cognitive Impairment: From vascular pathophysiology to potential clinical applications,” *Neuroscience*, vol. 579, pp. 267–283, Jul. 2025, doi: 10.1016/J.NEUROSCIENCE.2025.06.018.
- [28] L. Jiang, L. Xu, W. Sun, K. Bian, and Y. Wang, “Association between the coexistence of chronic kidney disease and sarcopenia with cardiovascular disease and mortality,” *Aging Clin. Exp. Res.*, vol. 37, no. 1, pp. 1–9, 2025, doi: 10.1007/s40520-025-03003-w.
- [29] Y. Tang *et al.*, “Comprehensive analysis of disease activity, neuropsychiatric symptoms, imaging features and risk factors for seizure in hospitalized patients with juvenile-onset neuropsychiatric systemic lupus erythematosus,” *Semin. Arthritis Rheum.*, vol. 74, no. August, p. 152798, 2025, doi: 10.1016/j.semarthrit.2025.152798.
- [30] D. F. Teshome, S. A. Balcha, T. A. Ayele, A. Atnafu, G. Mitike, and K. A. Gelaye, “Health extension workers led home-based multicomponent intervention improves linkage to hypertension care in northwest Ethiopia: cluster-randomized controlled trial,” *Trials*, vol. 26, no. 1, 2025, doi: 10.1186/s13063-025-08862-2.
- [31] J. E. Lee *et al.*, “Epidemiology and risk factors of fungal pathogens in sepsis: a prospective nationwide multicenter cohort study,” *BMC Infect. Dis.*, vol. 25, no. 1, 2025, doi: 10.1186/s12879-025-10722-y.
- [32] R. K. Garg, “Tuberculosis of the central nervous system: current concepts in diagnosis and treatment,” *Curr. Opin. Neurol.*, vol. 38, no. 4, pp. 406–413, Aug. 2025, doi: 10.1097/WCO.0000000000001358.
- [33] C. Geng *et al.*, “Artery fragment guided approach for enhancing cerebral aneurysm detection in TOF-MRA imaging,” *Comput. Methods Programs Biomed.*, vol. 269, p. 108906, Sep. 2025, doi: 10.1016/j.cmpb.2025.108906.
- [34] O. Waddell, A. McCombie, T. Glyn, J. Pearson, J. Keenan, and F. Frizelle, “The pathway to diagnosis of early-onset colorectal cancer: exploring diagnostic intervals and their effect on outcomes,” *Futur. Oncol.*, vol. 21, no. 19, pp. 2471–2482, Aug. 2025, doi: 10.1080/14796694.2025.2526319.

- [35] T. S. Tseng *et al.*, “Multi-level factors associated with low dose computed tomography lung cancer screening in the United States,” *J. Natl. Med. Assoc.*, vol. 117, no. 4, pp. 225–234, 2025, doi: 10.1016/j.jnma.2025.05.002.
- [36] N. Reimes *et al.*, “The patient journey of head neck cancer from primary care to diagnosis – A retrospective cohort study in the Netherlands,” *Cancer Epidemiol.*, vol. 97, p. 102865, Aug. 2025, doi: 10.1016/j.canep.2025.102865.
- [37] M. H. Tsai, J. B. Vo, J. X. Moore, C. Ramin, A. Guha, and Y. Dong, “Mediating effects of mental and physical health on the association between chronic disease conditions and colorectal cancer screening utilization among breast cancer survivors,” *PLoS One*, vol. 20, no. 8 August, pp. 1–12, 2025, doi: 10.1371/journal.pone.0328353.
- [38] A. Askarinejad, D. A. Lane, P. Sadeghipour, M. Haghjoo, and G. Y. H. Lip, “Stroke prevention in atrial fibrillation: A narrative review of current evidence and emerging strategies,” *Eur. J. Clin. Invest.*, no. May, pp. 1–26, 2025, doi: 10.1111/eci.70082.
- [39] Z. Mobini *et al.*, “Risk Adjusted Continuous Monitoring of Postoperative Mortality After Cardiac Surgery,” *Health Serv. Res.*, vol. 60, no. 4, Aug. 2025, doi: 10.1111/1475-6773.14607.
- [40] S. Alp *et al.*, “Joint transformer architecture in brain 3D MRI classification : its application in Alzheimer ’ s disease classification,” *Sci. Rep.*, pp. 1–12, 2024, doi: 10.1038/s41598-024-59578-3.