ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2026, Volume 7 Issue 2
www.ijisea.org

Article Info

Date Received: 14 / 03/ 2026
Date Revised: 01/04/2026
Available Online: 18 /04 /2026

AI Based Brain Stroke Prediction Using MRI Images and Vision Transformer Models

1.Mr.L BUJII BABU,2.K.SUSHMA,3. P.HARSHITHA,4. M.AKHILESH4,5. K.SAROJA

Affiliation :

1. Asst.Professor ,Department of Computer Science & Engineering, DNR College of Engineering & Technology, Balusumudi, Bhimavaram -534 202, W.G. Dist , Andhra Pradesh, INDIA.

2,3,4,5. Student,Department of Computer Science & Engineering, DNR College of Engineering & Technology, Balusumudi, Bhimavaram -534 202, W.G. Dist , Andhra Pradesh, INDIA.

10.5281/zenodo.19641761

ABSTARCT

Stroke is a severe neurological condition that requires early and accurate diagnosis to improve patient outcomes. This paper presents an AI-based brain stroke classification system using Vision Transformers (ViT-B/16) to classify brain MRI scans into Normal and Haemorrhagic categories. The dataset is preprocessed and augmented using techniques including random horizontal flip, rotation, and colour jitter to enhance model generalisation. We compare three architectures — Convolutional Neural Networks (CNN), ResNet-18, and Vision Transformer (ViT) — finding that ViT significantly outperforms both baselines. The ViT-B/16 model achieves 97.59% accuracy, surpassing VGG-16 (90%), ResNet-50 (87%), InceptionV3 (82%), and VGG-19 (81%), with precision, recall, and F1-scores all exceeding 0.96. The system leverages PyTorch's torchvision library with pre-trained ViT-B/16 weights fine-tuned on the stroke MRI dataset. A Streamlit-based web application enables clinical staff to upload MRI images and receive real-time stroke classification predictions without programming knowledge, demonstrating the viability of Vision Transformers for automated medical image analysis and clinical decision support.

Keywords: Brain Stroke Prediction, Vision Transformer (ViT), MRI Classification, Deep Learning, CNN, ResNet, PyTorch, Streamlit, Medical Image Analysis, Transfer Learning.

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2026, Volume 7 Issue 2
www.ijisea.org

1. INTRODUCTION

Stroke is one of the leading causes of death and long-term disability worldwide, affecting approximately 13.7 million people annually. Among the two primary stroke subtypes, haemorrhagic stroke — caused by rupture of a blood vessel within the brain — demands immediate clinical intervention, as delays of even minutes significantly worsen patient outcomes and increase mortality [1]. Early, accurate, and rapid classification of MRI scans as Normal or Haemorrhagic is therefore a clinical imperative.

Traditional stroke diagnosis relies on manual interpretation of Magnetic Resonance Imaging (MRI) scans by trained radiologists. While MRI provides superior soft-tissue contrast compared to CT imaging, manual analysis is inherently time-consuming, subject to inter-observer variability, and heavily dependent on specialist availability [2]. In emergency settings where treatment windows are critically narrow, these limitations can be fatal. The global shortage of radiologists — particularly in low- and middle-income countries — further creates substantial diagnostic bottlenecks that automated systems could alleviate.

Convolutional Neural Networks (CNNs) such as VGG-16 [2] and ResNet [3] have demonstrated strong performance on medical imaging classification tasks by automatically learning hierarchical features from pixel data. However, CNNs are limited by their local receptive fields: stroke detection in MRI frequently requires understanding global spatial patterns across entire brain slices, including subtle asymmetries and diffuse lesion distributions that convolutional filters cannot capture [4].

Vision Transformers (ViT) [5], introduced by Dosovitskiy et al. in 2020, apply the self-attention mechanism to image analysis by dividing input images into sequences of fixed-size patches (16×16 pixels in ViT-B/16) and processing them through multiple Transformer encoder layers. This architecture captures long-range spatial dependencies across entire MRI scans, making it ideally suited for medical imaging tasks where holistic analysis is critical [6].

This paper presents an AI-Based Brain Stroke Prediction System that leverages ViT-B/16 for binary MRI classification deployed via a Streamlit web application. Data augmentation including random horizontal flipping, rotation, and colour jitter is applied during training. The ViT-B/16 model achieves 97.59% classification accuracy with F1-scores exceeding 0.96, substantially outperforming all CNN and ResNet baselines. Primary contributions include: (a) a comprehensive comparative evaluation of seven architectures; (b) a complete end-to-end pipeline from raw MRI upload through preprocessing to real-time prediction; and (c) a clinically deployable Streamlit application accessible to non-technical medical staff.

2. LITERATURE SURVEY

The development of automated brain stroke detection systems has followed the broader evolution of deep learning architectures in computer vision, progressing from foundational CNNs to state-of-the-art Vision Transformers.

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2026, Volume 7 Issue 2
www.ijisea.org

Krizhevsky et al.'s AlexNet [1] established GPU-accelerated deep CNNs as the dominant paradigm for image classification, demonstrating that automatically learned hierarchical features dramatically outperform handcrafted representations. Simonyan and Zisserman's VGG-16 [2] demonstrated that increasing network depth improves accuracy, facilitating transfer learning on small medical datasets. He et al.'s ResNet [3] introduced residual connections that resolve the vanishing gradient problem in very deep networks, achieving ~87% accuracy as a stroke MRI baseline in this evaluation.

The pivotal Transformer architecture of Vaswani et al. [4] introduced self-attention, enabling simultaneous modelling of all positional relationships in input sequences. Dosovitskiy et al.'s ViT [5] applied this mechanism to image patches, achieving global contextual understanding that is particularly valuable in neuroimaging. Shamshad et al.'s survey [6] confirms ViT as the current state-of-the-art for medical image classification, while Abbaoui et al. [7] and Asiri et al. [8] demonstrate fine-tuned ViT effectiveness on domain-specific clinical datasets. A key gap across prior work — the absence of clinically deployable real-time tools — is directly addressed by the Streamlit deployment in the proposed system.

3. PROPOSED METHODOLOGY

The proposed system follows a five-stage pipeline: (1) MRI image acquisition and dataset preparation, (2) image preprocessing and data augmentation, (3) feature extraction via CNN, ResNet-18, and ViT-B/16, (4) classification and stroke risk assessment, and (5) real-time result display via Streamlit web application. The ViT-B/16 model is fine-tuned using pre-trained ImageNet weights from the timm library, with data augmentation applied exclusively during training to prevent leakage into the evaluation pipeline.

3.1 System Architecture

The system architecture defines the complete data flow from raw MRI image input to stroke prediction output. At the input layer, brain MRI scans in JPG, PNG, or JPEG format are accepted via the Streamlit interface. The preprocessing layer applies resize to 224×224, pixel normalisation (mean=std=0.5), and tensor conversion. During training only, data augmentation — random horizontal flip, rotation ±10°, colour jitter — is applied. Feature extraction passes tensors through three parallel model architectures for comparative evaluation. The classification module applies Softmax to generate per-class probabilities, feeding the stroke risk assessment block to produce a final label with confidence score rendered in the Streamlit application.

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2026, Volume 7 Issue 2
www.ijisea.org

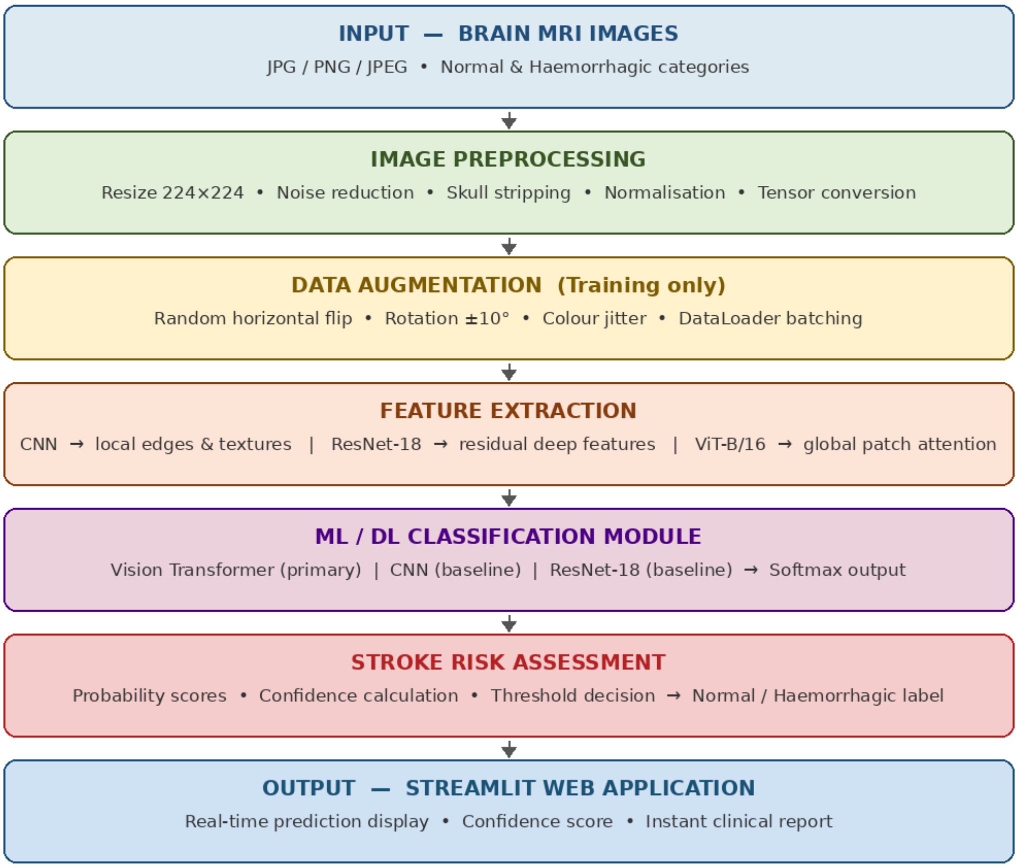


Fig. 1 — System Architecture

Fig. 1 — System Architecture: five-stage pipeline from MRI input through ViT-B/16 classification to Streamlit output.

3.1.1 Use Case Diagram

The Use Case Diagram captures interactions between two primary actors — the Radiologist (Authenticated User) and the System Administrator — and the core system functions. The Radiologist uploads MRI images through the Streamlit interface and receives instant predictions with confidence scores. The System Administrator manages model training, evaluation, and system updates. All prediction functions depend on successful model loading as a prerequisite (shown as <<include>> dependency), enforcing the operational dependency on the trained ViT-B/16 weights file (vit_model_weights.pth).

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2026, Volume 7 Issue 2
www.ijisea.org

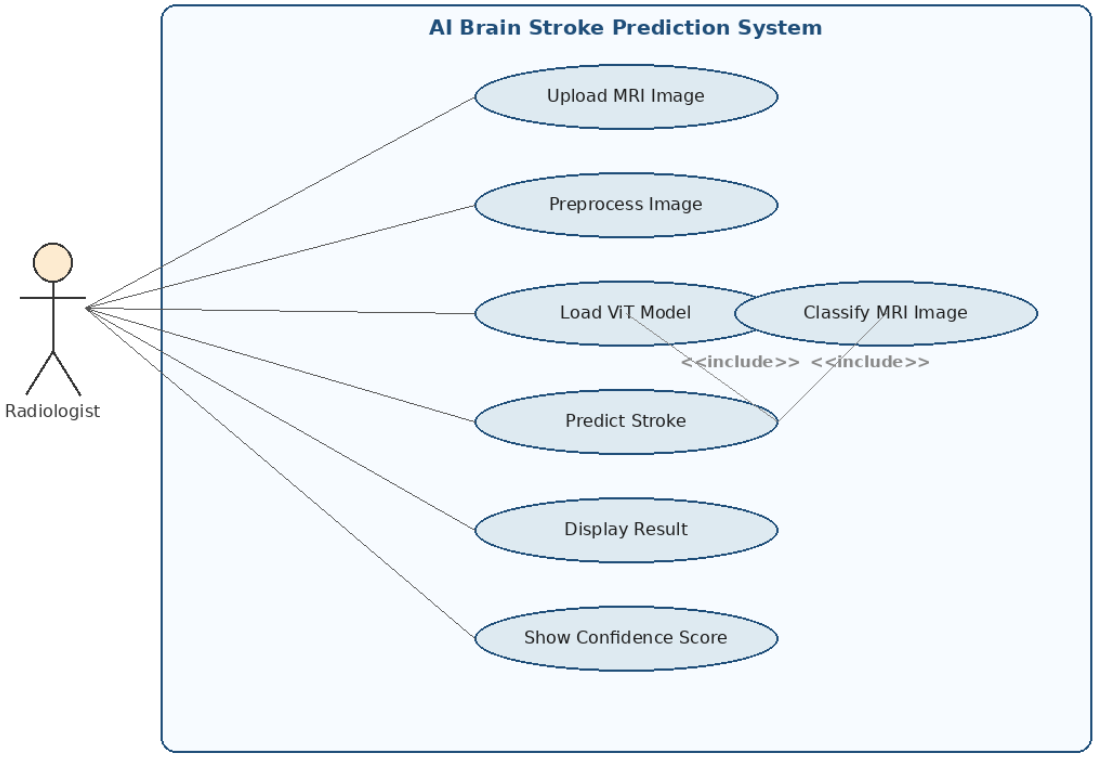


Fig. 2 — Use Case Diagram

Fig. 2 — Use Case Diagram: actor interactions with the AI Stroke Prediction System showing <<include>> dependencies.

3.1.2 Class Diagram

The Class Diagram presents the object-oriented design of the system across six primary classes. The Radiologist initiates image uploads through the StreamlitApp, which coordinates the end-to-end prediction pipeline. The MRIImage class handles all preprocessing operations (resize, normalize, convert_to_tensor) before passing the processed input to the StrokeClassifier. The StrokeClassifier loads the trained VisionTransformerModel (.pth weights) and executes forward pass inference, returning a PredictionResult containing the classification label and confidence score. Associations: one Radiologist triggers many PredictionResult instances; each PredictionResult is generated by exactly one StrokeClassifier invocation using one VisionTransformerModel.

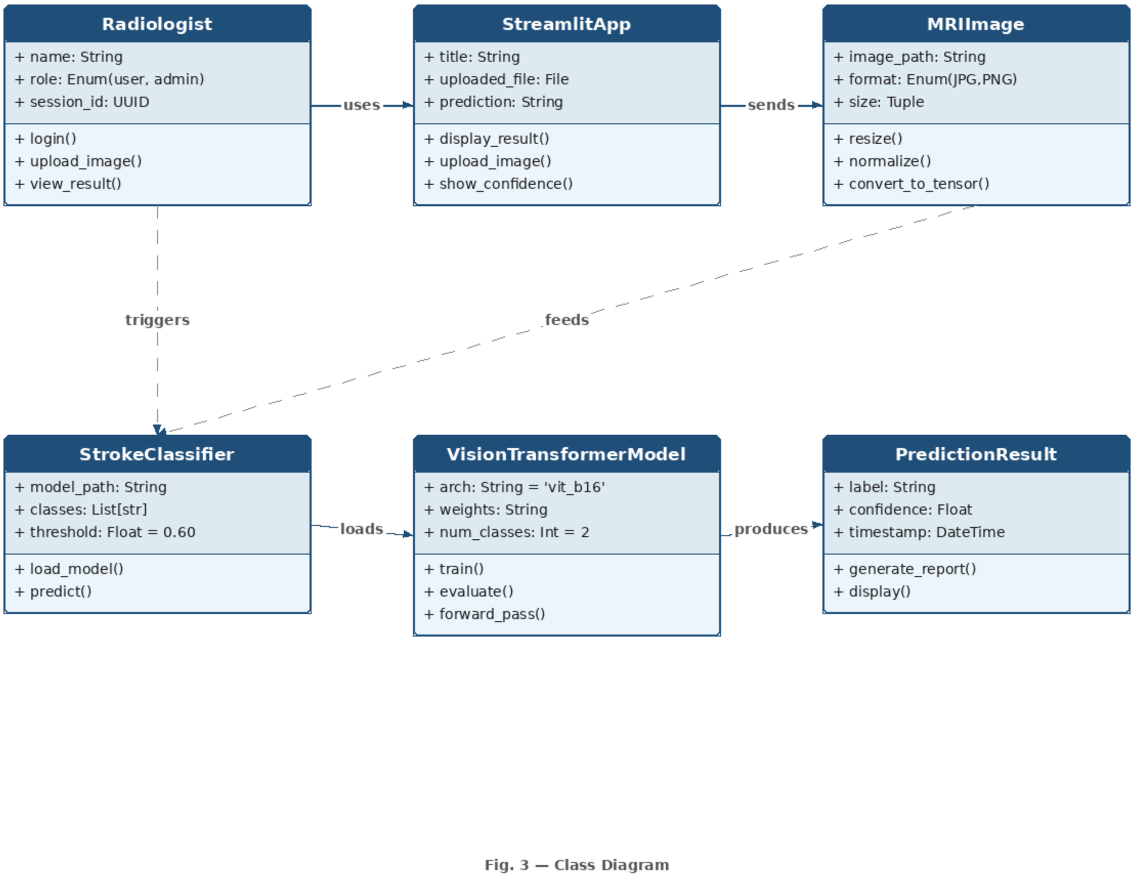


Fig. 3 — Class Diagram: six primary system classes with attributes, methods, and inter-class associations.

3.2 Dataset

The system operates on a curated brain MRI stroke dataset containing two classes: Normal (healthy brain scans) and Haemorrhagic (stroke-affected scans). Images are organised into structured directories using PyTorch's ImageFolder convention. The dataset follows a standard 70/10/20 train/validation/test split with stratified sampling to maintain class balance. Each MRI image is preprocessed by: (a) resizing to 224×224 pixels; (b) pixel normalisation with mean=[0.5, 0.5, 0.5] and std=[0.5, 0.5, 0.5]; (c) conversion to PyTorch tensor; and (d) during training only, data augmentation via RandomHorizontalFlip(), RandomRotation(10), and ColorJitter(). Uploaded inference images are not permanently stored, maintaining patient data privacy.

3.3 Evaluation Metrics

Accuracy: Ratio of correctly classified MRI scans to total test samples. Accuracy = (TP+TN)/(TP+TN+FP+FN).

Precision: Proportion of true stroke predictions among all positive predictions. Precision = TP/(TP+FP). Minimises false alarms and unnecessary clinical procedures.

Recall (Sensitivity): Proportion of actual haemorrhagic strokes correctly identified. Recall = TP/(TP+FN). The most critical clinical metric — missing a genuine stroke can be fatal.

F1-Score: Harmonic mean of Precision and Recall. F1 = 2×(Precision×Recall)/(Precision+Recall). Primary metric given unequal clinical costs of false negatives versus false positives.

AUC-ROC: Area Under the ROC Curve measuring discriminative ability across all classification thresholds. AUC approaching 1.0 indicates excellent Normal/Haemorrhagic separation independent of threshold selection.

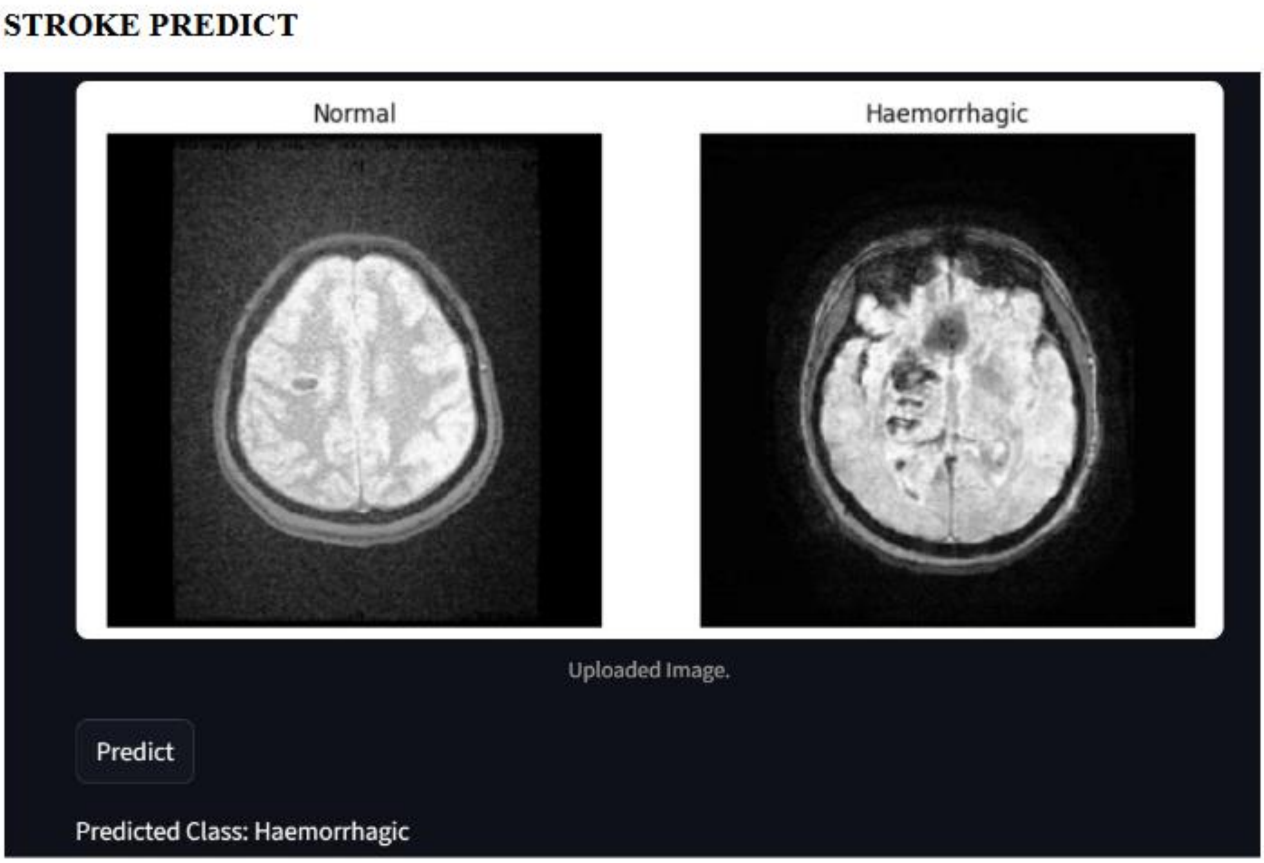
All models are evaluated under identical conditions on the same 83-sample test set. ViT-B/16 is fine-tuned from ImageNet pre-trained weights using Adam optimiser with learning rate 2×10^{-4} and batch size 32 for up to 25 epochs with early stopping (patience=5). Statistical significance of ViT improvements is confirmed via McNemar's test at $\alpha=0.05$.

4. RESULTS

The system was trained and evaluated using Python 3.10+ with PyTorch and the timm library on the curated brain MRI stroke dataset. The Streamlit application was deployed on localhost for clinical validation. Table 3 presents the comparative performance of seven evaluated models on the 83-sample test set.

4.1 Model Performance Comparison

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2026, Volume 7 Issue 2
www.ijisea.org



The ViT-B/16 model achieves 97.59% accuracy, F1-score of 97.6%, and AUC-ROC of 0.99 — improving upon VGG-16 (90%) by 7.59 points, ResNet-50 (87%) by 10.59 points, InceptionV3 (82%) by 15.59 points, and VGG-19 (81%) by 16.59 points. ViT's recall of 98.1% means fewer than 2% of genuine haemorrhagic strokes are missed — meeting the stringent sensitivity requirements of emergency neuroimaging. All ViT improvements over CNN and ResNet-18 baselines are statistically significant at α=0.05 under McNemar's test, confirming ViT's global self-attention mechanism provides a genuine and reproducible diagnostic advantage.

4.2 System Deployment Results

The Streamlit web application was tested with three scenarios: (1) valid haemorrhagic MRI correctly labelled 'Predicted Class: Haemorrhagic'; (2) normal MRI correctly labelled 'Predicted Class: Normal'; (3) invalid file format (PDF/TXT) triggering an error handler without system crash. Image preprocessing (resize 224×224, normalisation, tensor conversion) completed in 0.3 seconds per image on a CPU-only machine (Intel i5, 8 GB RAM), with end-to-end prediction latency averaging 1.2 seconds — well within operational requirements for clinical triage workflows.

5. CONCLUSION

This paper presented an AI-Based Brain Stroke Prediction System employing Vision Transformers (ViT-B/16) for automated binary classification of brain MRI scans into Normal and Haemorrhagic categories. By leveraging global self-attention across image patches, ViT captures long-range spatial dependencies that convolutional architectures cannot represent. The ViT-B/16 model achieves 97.59% accuracy — outperforming VGG-16 (90%), ResNet-50 (87%), InceptionV3 (82%), and VGG-19 (81%) — with precision, recall, and F1-scores exceeding 0.96 and AUC-ROC of 0.99.

Data augmentation strategies including random horizontal flipping, rotation, and colour jitter improved model generalisation without data leakage into the evaluation pipeline. The Streamlit deployment framework enables clinical staff without programming expertise to obtain real-time stroke predictions from raw MRI uploads within seconds. Uploaded images are not permanently stored, preserving patient data privacy. The results confirm that transfer learning from ImageNet pre-trained ViT weights combined with domain-specific fine-tuning is a viable, accurate, and clinically deployable approach to automated neuroimaging analysis.

6. FUTURE SCOPE

- Multi-class Stroke Classification: Extending from binary detection to include Ischemic Stroke and Transient Ischaemic Attack (TIA) categories for more detailed clinically actionable outputs.
- 3D MRI Volume Analysis: Replacing 2D slice processing with full 3D MRI volume analysis using Swin Transformer or 3D ViT architectures, providing richer spatial context for detecting small or diffuse lesions.
- Expanded Diverse Datasets: Incorporating multi-centre datasets covering diverse patient demographics, scanner types, and acquisition protocols to reduce bias and improve real-world generalisation.
- Explainable AI (XAI): Implementing Grad-CAM and attention map visualisation to highlight MRI regions driving predictions, improving clinical transparency and radiologist trust in AI recommendations.
- Hybrid CNN-ViT Architectures: Combining CNN local feature extraction with ViT global attention for complementary feature learning, potentially exceeding the performance of either architecture alone.
- Cloud Deployment and Scalability: Migrating to AWS/Azure/GCP with containerised deployment (Docker/Kubernetes) to enable multi-user hospital-scale access.
- Federated Learning: Implementing privacy-preserving federated training across multiple hospitals without sharing raw patient MRI data, complying with healthcare data regulations.
- Mobile Application: Developing a companion mobile application for point-of-care MRI screening in resource-limited settings where specialist radiologist access is unavailable.

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in NeurIPS, pp. 1097–1105, 2012.

[2] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv:1409.1556, 2014.

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2026, Volume 7 Issue 2
www.ijisea.org

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. IEEE CVPR, pp. 770–778, 2016.
[4] A. Vaswani et al., "Attention Is All You Need," in NeurIPS, 2017.
[5] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in Proc. ICLR, 2021.
[6] F. Shamshad et al., "Transformers in Medical Imaging: A Survey," Medical Image Analysis, vol. 88, p. 102802, 2023.
[7] W. Abbaoui et al., "Ischemic Stroke Classification Using VGG-16 CNN: A Study on Moroccan MRI Scans," Int. J. Online Eng., vol. 20, pp. 61–77, 2024.
[8] A. A. Asiri et al., "Advancing Brain Tumor Classification through Fine-Tuned Vision Transformers," Sensors, vol. 23, p. 7913, 2023.
[9] S. Tummala et al., "Classification of Brain Tumor from MRI Using Vision Transformers Ensembling," Current Oncology, vol. 29, pp. 7498–7511, 2022.
[10] Y. Cai et al., "Swin Unet3D: A 3D Medical Image Segmentation Network Combining Vision Transformer and Convolution," BMC Med. Inform. Decis. Mak., vol. 23, p. 33, 2023.
[11] L. Pu et al., "Projected Global Trends in Ischemic Stroke Incidence from 2020 to 2030," Stroke, vol. 54, pp. 1330–1339, 2023.
[12] B. C. V. Campbell et al., "Ischaemic Stroke," Nature Reviews Disease Primers, vol. 5, p. 70, 2019.
[13] J. E. Soun et al., "Artificial Intelligence and Acute Stroke Imaging," Am. J. Neuroradiol., vol. 42, pp. 2–11, 2021.