Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

Enhanced ResNet-Vision Transformer using Optimization Techniques for Brain Stroke Detection

Jacophine Susmi S.[1*], Gunavathie M.A.[2], Kavya S.[3], Priya J.[4]

[1]Assistant Professor, University College of Engineering Tindivanam, Melpakkam, India.
[2]Assistant Professor, Department of Information Technology, Easwari Engineering College, Ramapuram, Chennai, India.
[3]Assistant Professor, Department of Computer Science and Engineering (Cyber Security), Easwari Engineering College, Ramapuram, Chennai, India.
[4]Assistant Professor, Department of Artificial Intelligence and Data Science, Easwari Engineering College, Ramapuram, India.

E-mail: [1*]jacophine.susmi@gmail.com, [2]gunavathie.ap@gmail.com, [3]kavyasabapathy@gmail.com, [4]pria.manikandan@gmail.com

Orcid ID: [1*]0009-0001-2079-6734, [2]0000-0002-0767-779X, [3]0009-0004-3962-5072, [4]0000-0002-1868-3917

Abstract

Early recognition of brain stroke is the key to prompt clinical management and better outcomes. Accurate identification of stroke-affected tissue from normal brain tissue structures in Computed Tomography (CT) images, however, is still a difficult task because of its subtle intensity differences and complex anatomical characteristics. In this regard, this paper proposes an Enhanced Black-winged Kite Optimized ResNet–Vision Transformer Network (HBK-ResViTNet) for automated brain stroke detection. The first step in the CT processing includes detailed noise removal, image normalization, skull stripping, and contrast enhancement in order to improve the image quality and highlight regions of interest for diagnostic purposes. Then, ResNet50 is used to generate a discriminative local spatial representation, and Vision Transformer (ViT) is used to understand the long-range contextual information and achieve global feature interaction. The complementary features generated by these two networks are effectively merged to acquire a powerful feature representation for stroke identification. Additionally, the Enhanced Black-winged Kite Optimization (EBKO) algorithm is used to tune the network hyperparameters for better convergence speed and classification accuracy. Finally, the optimized feature representations are fed into fully connected layers to classify CT images into multiple classes. The proposed framework is evaluated using the RSNA Intracranial Hemorrhage Detection Dataset, a large-scale, multi-institutional brain CT dataset containing expert annotations for five major intracranial haemorrhage subtypes. The experimental outcomes show that it attains higher accuracy, precision, recall, and F1-score than traditional deep learning techniques. The results suggest that the suggested HBK-ResViTNet provides a dependable, efficient, and scalable computer-aided system for early diagnosis of brain stroke, which would be useful in clinical decision-making and management of stroke patients.

Keywords: Stroke Detection, Hybrid Deep Learning, ResNet50, Vision Transformer (ViT),

*** Corresponding Author**

Journal of Innovative Image Processing, December 2026, Volume 8, Issue 4, Pages 1599-1619
DOI: https://doi.org/10.36548/jiip.2026.4.013
Received: 04.09.2026, received in revised form: 08.10.2026, accepted: 31.10.2026, published: 10.11.2026
© 2026 Inventive Research Organization. This is an open access article under the Creative Commons Attribution-NonCommercial International (CC BY-NC 4.0) License

Enhanced Black-winged Kite Optimization (EBKO).

1. Introduction

Stroke is a life-threatening brain disease caused by a lack of blood supply to the brain and can cause irreversible damage if not diagnosed early. Early and precise diagnosis is crucial to lowering the death rate and long-term disabilities [1]. Computed Tomography (CT) scans are commonly used in emergency rooms for fast and efficient detection of brain abnormalities. But early signs of stroke can be subtle and blend in with normal brain structures. Thus, manual diagnosis is highly dependent on radiologist experience, which may lead to variability in diagnosis [2]. These enable the automated extraction of features without manual feature engineering feature design and enhance diagnostic reproducibility. Specifically, convolutional models have been widely used to analyse CT images as they can learn spatial features. Effective medical image analysis requires the extraction of both local spatial features and global contextual information to accurately characterize complex pathological patterns. Hybrid deep learning models have recently been proposed to meet this need by hybridizing feature learning mechanisms [3].

However, while great progress has been made in deep learning-based classification of medical images, early detection of stroke is still difficult due to subtle intensity differences in CT images [4]. This makes it difficult to detect non-local or faint stroke patterns. Models based on Transformers capture long-range dependencies but rely on large-scale data, which is not often available in medical applications. Further, most existing approaches use either fixed or user-defined hyperparameters, resulting in suboptimal model performance and convergence behaviours. Class imbalance and scarcity of labeled data also limit generalization in medical data [5]. The absence of an integrated approach to jointly optimize feature extraction and model settings is another problem. Traditional training approaches don't update the learning parameters during training. This causes a higher computational burden and unstable classification. Therefore, there is a need for an adaptive hybrid metaheuristic-optimized model [6].

Conventional stroke detection techniques are mainly based on manual feature engineering feature extraction techniques, including texture descriptors, histogram features, and intensity statistics. These features are often divided by conventional machine learning classifiers, like Support Vector Machines (SVM), k-Nearest Neighbours (kNN),in addition Random Forest (RF) [7]. These approaches offer primitive classification performance, but rely deeply on feature engineering and expert knowledge [8]. The advent of deep learning has revolutionised feature learning from medical images with convolutional neural networks (CNNs), such as AlexNet, VGG, and ResNet [9]. Recently, Vision Transformers have been applied to medical images because they have the ability to capture long-range relationships using self-attention [10]. Moreover, the current methods do not use smart optimization for hyperparameter tuning and therefore do not perform optimally. Metaheuristic techniques like Genetic Algorithms and Particle Swarm Optimization have been used, but they tend to converge prematurely and are unstable. Ensemble models ensure robust detection, but are computationally expensive. The main contribution of the proposed approach is listed below;

- The HBK-ResViTNet model is proposed, which is a hybrid deep learning approach where ResNet50 is used for local feature extraction and ViT for global feature modeling to detect stroke in brain CT images.

- A new hyperparameter optimization approach based on EBKO, as well as an attacking, migrating, and Lévy flight strategy is developed for optimizing the network architecture.
- The purpose of designing an integrated hybrid deep learning framework is to achieve improved accuracy and generalizability for stroke detection compared to traditional approaches.

2. Related Works

The progress in automated identification of stroke from brain CT images has come a long way with the development of machine learning and deep learning techniques. In general, the literature suggests a more comprehensive hybrid approach that considers effective local-global feature learning and smart optimization techniques to achieve better accuracy in stroke detection.

Table 1. Comparative validation of existing deep learning-based stroke detection methods in brain CT/MRI images

Reference	Method	Strengths	Limitations
Ahmed et al., [11]	Hybrid Swin Transformer–BiLSTM	Effectively integrates spatial and temporal representations using hybrid attention and sequence modeling, leading to improved feature richness in multimodal stroke analysis.	Computationally expensive due to the combined transformer and recurrent architecture, and requires high training resources
Mostafapour et al., [12]	CNN + XGBoost	Combines deep feature extraction with a gradient boosting classifier, improving classification robustness on structured feature spaces	Lacks effective modeling of global spatial dependencies and depends heavily on manual feature engineering feature transformation stages
Mandhare et al., [13]	Kronecker Network with Metaheuristic Optimization	Enhances learning performance through optimized deep network parameter search, improving convergence behaviour in CT-based stroke detection	The optimization process increases computational complexity and requires careful parameter tuning for stable training.
Shangguan et al., [14]	Progressive Feature Fusion Network	Utilizes a hierarchical feature fusion strategy that improves lesion boundary representation and segmentation consistency	High architectural complexity and increased inference time due to multi-stage fusion design
Hossain et al., [15]	ViT–LSTM Hybrid Model	Captures both global contextual dependencies and sequential feature correlations, improving classification capability	Performance depends heavily on large-scale training data and exhibits high computational demand.
Thakre et al., [16]	Hybrid Deep Learning Model	Enhances model robustness by integrating multiple deep learning components for improved feature representation	Limited interpretability and lacks adaptive optimization of network parameters.
Sabir et al., [17]	CNN-based Stroke Detection	Efficient in extracting local spatial features and suitable for early-stage stroke pattern recognition	Unable to effectively model long-range dependencies and global contextual relationships
Abumihsan et al., [18]	Attention-based Feature Fusion (CBAM)	Improves discriminative feature learning through channel and spatial attention mechanisms	Does not include adaptive hyperparameter optimization, leading to suboptimal performance in varying datasets

Table 1 outlines recently published works in the literature on brain stroke detection using CT and MRI images [11]-[18], detailing their approaches, advantages, and disadvantages. Ahmed et al. [11] utilized a hybrid Swin Transformer and BiLSTM approach to enhance multimodal feature learning, though it is computationally expensive. Mostafapour et al. [12] employed a CNN with XGBoost to improve classification accuracy, but it lacks the capture of global contextual information. Mandhare et al. [13] proposed a deep Kronecker network with a metaheuristic optimizer that enhances convergence, but adds computational cost. Shangguan et al. [14] proposed a progressive feature fusion network that improves lesion segmentation, but it has a complicated multi-step approach. Hossain et al. [15] developed a ViT-LSTM architetcure that efficiently captures long-range dependencies, but requires large datasets. Thakre et al. [16] proposed a hybrid deep learning architecture that enhances robustness, but it is not interpretable and lacks hyperparameter tuning. Sabir et al. [17] adopted a CNN-based technique to extract local information, but struggled to learn global information. Abumihsan et al. [18] presented an attention-based feature fusion mechanism that enhances feature discriminability, but it doesn't involve an optimization mechanism, making it rigid. In summary, the above studies reveal that current approaches either concentrate on local or global feature extraction and do not include an optimization module, which justifies the design of the proposed HBK-ResViTNet.

3. Proposed System Model: HBK-ResViTNet Framework

The proposed HBK-ResViTNet is a powerful and scalable deep learning-based framework for early identification of stroke in brain CT scans. The model starts with a well-designed preprocessing workflow to overcome the issues of medical images, including noise, poor contrast, and clutter. It begins with Gaussian filtering to reduce high-frequency noise while maintaining crucial structures. Next, intensity normalization is applied to standardise the distribution of pixel values in the dataset, enhancing the model's generalizability. Subsequently, skull stripping is performed to exclude non-brain structures (skull, scalp, etc.) and artifacts to confine the model's attention to the brain. Finally, CLAHE is applied to enhance subtle intensity differences, highlighting early signs of stroke lesions. Next preprocessing, the enhanced images are processed by a feature extraction component based on ResNet50, which extracts local spatial features such as edges, textures, and micro-scale features associated with stroke, such as ischemic and haemorrhagic lesions, using deep residual learning. These residual connections help to avoid the vanishing gradient problem and allow the extraction of high-level discriminative features from complex medical images. The entire process of the proposed HBK-ResViTNet is presented in Figure 1, where the CT images are pre-processed, followed by hybrid feature extraction from ResNet50 and ViT, then optimised by the EBKO algorithm, and finally classified.

The proposed HBK-ResViTNet framework fuses ResNet50, ViT, and EBKO to leverage their complementary capabilities. ResNet50 effectively extracts local spatial features such as texture variations, lesion boundaries, and subtle intensity changes related to stroke lesions. However, convolutional networks often struggle to learn long-range context between remote brain regions. To address this, a Vision Transformer module is integrated to capture global dependencies using a self-attention mechanism. Moreover, hyperparameter tuning is crucial for the performance of deep learning architectures. To optimize this, EBKO is employed to automatically tune important hyperparameters, leading to faster convergence, improved classification performance, and stability of models. Thus, the proposed framework

simultaneously exploits local feature learning, global contextual modelling, and adaptive optimization for accurate and precise early stroke detection.



Figure 1. Block diagram of the proposed HBK-ResViTNet framework with Enhanced Black-winged Kite Optimization (EBKO) for automated brain CT image classification

3.1 Image Preprocessing – Multi-Stage Enhancement Pipeline

CT scans of the brain are prone to noise, intensity variations, and extraneous anatomical features, which may mask subtle stroke patterns and affect model performance. To address these issues, a systematic preprocessing approach can be employed to improve image quality and focus on relevant areas for diagnosis. This is formulated as

$$I_g(x,y) = I(x,y) * G(x,y) \tag{1}$$

in which $I(x,y)$is the input CT image, $G(x,y)$is the Gaussian kernel, *is the convolution operator,in addition$I_g(x,y)$is the filtered CT image. The transformed image is

$$I_n = \frac{I_g - \mu}{\sigma} \tag{2}$$

where I_g is the denoised image, μ is the mean intensity, σ is the standard deviation, and $I_n$is the normalized image. This involves masking the brain region with a binary mask and is formulated as

$$I_s(x,y) = I_n(x,y) \cdot M(x,y) \tag{3}$$

where $M(x,y)$is the binary mask for the brain (1 for brain; 0 otherwise) and $I_s(x,y)$is the skull-stripped image. The enhanced image is given by

$$I_e = \text{CLAHE}\,(I_s) \tag{4}$$

where $I_s$is the skull-stripped image, and $I_e$is the output enhanced image used for further processing. These enhancements ensure the input to the hybrid deep learning model is consistent and rich in information, leading to improved early detection of stroke.

3.2 Local Feature Extraction – ResNet50 Backbone

After preprocessing, the refined CT images are processed by a ResNet50 backbone to learn local spatial features that are important for the early detection of stroke. ResNet50 is a deep 50-layer convolutional network composed of stacked bottleneck residual blocks formed by a 1×1 - 3×3 - 1×1 pattern of convolutional operations. These operations reduce channels for dimensionality, learn spatial patterns, and increase dimensions, respectively. This architecture allows for effective deep feature extraction with reduced complexity, making it ideal for medical image processing. The residual mapping is defined as

$$y = F(x, W) + x \tag{5}$$

Here, $x \in \mathbb{R}^{Htimes Wtimes C}$represents the input feature map with spatial dimensions H, W, and C channels, $F(x, W)$is the residual function learned by stacked convolutional layers with weightson the output feature map.

3.3 Global Context Modeling – Vision Transformer (ViT)

To collect both the local and the global information of brain CT images, a ViT module is used to capture long-range correlations and contextual information. Suppose an image $I \in \mathbb{R}^{H\times W\times C}$is resized into N patches of $P \times P$. The patches are linearly projected into a latent space and constitute a sequence of tokens for a transformer. The key component of ViT is the multi-head self-attention, which allows ViT to learn interactions between all patches. The self-attention mechanism can be given as

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{6}$$

Here, Q, K, and V define the query, key, and value measures computed from the input embeddings, and $d_k$is the length of the key vectors.Algorithm 1 presents the Pseudocode of HBK-ResViTNet with EBKO.

The extracted local features from ResNet50 are combined with the global contextual features extracted from the Vision Transformer through a feature fusion process to obtain a global feature representation. The local features from ResNet50 focus on fine-grained structural details like lesion boundaries, texture patterns, and intensity variations, while the ViT module is adept at learning long-range spatial relationships and context across various brain regions. These complementary feature representations are fused together and used as a single feature vector, allowing the model to make use of both local and global information. The fused feature representation is then passed to the classification layer, where the final stroke prediction is calculated. By combining these features, the model's discriminative ability is strengthened, leading to better classification performance and robustness.

Once the feature extraction step is done, the extracted local features from ResNet50

and the generated global contextual features from the Vision Transformer are fused to create a single feature representation. The ResNet50 module is able to extract fine-grained structural and texture information, while the ViT module is able to extract long-range spatial and contextual information across different brain regions. These auxiliary features are combined into a single feature vector, and the model exploits both local and global information. The fused representation is then sent to the fully connected classification layer for the final stroke identification, which will increase the discriminative power of features and improve the classification accuracy.

3.4 Hyperparameter Optimization – Enhanced Black-winged Kite Optimization (EBKO)

In contrast with traditional optimization algorithms like PSO, GA, GWO, and ACO, the proposed EBKO algorithm unites attacking behavior, migration behavior, and Lévy flight-based exploration into a single optimization framework. The attacking mechanism helps improve local exploitation around the best solutions, while the migration mechanism ensures good global exploration by steering the candidate solutions to high-quality regions of the search space. Furthermore, the Lévy flight strategy incorporates adaptive long-distance movements that enable the optimizer to avoid local optima and maintain population diversity. These characteristics help EBKO converge more quickly and can help select the hyperparameters of the proposed HBK-ResViTNet model [19].

3.4.1 Initialization

The process of finding optimal hyperparameter settings starts with the creation of a population of solutions, with each solution representing a particular hyperparameter configuration of HBK-ResViTNet. These parameters are sampled between reasonable bounds, based on empirical observations (e.g., learning rate, batch size, attention heads). This process guarantees that shallow and deep networks are considered early in the search, Which helps to prevent early stagnation at local optima. Initialization is defined as:

$$X_i = BK_{lb} + rand \cdot (BK_{ub} - BK_{lb}) \tag{7}$$

where $X_i$is the i-th hyperparameter vector, BK_{lb} and BK_{ub} are the lower and upper bounds, and $mathrmrand \in [0, 1]$. This can be the initial population for optimization.

3.4.2 Fitness Evaluation

After initialization, the fitness of each solution is assessed using a fitness function determined by the HBK-ResViTNet performance. For each set of hyperparameters, the architecture is trained (or partially trained) on the CT dataset and validated to assess performance. The fitness function ensures the evolutionary process aligns line with the detection goal and is defined as:

$$F_i = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

where TP, TN, FP, and FN are true positives, true negatives, false positives, and false negatives, respectively. This is used to guide the algorithm to optimal hyperparameters

3.4.3 Attacking Behavior (Local Exploitation)

The attacking phase optimises solutions by concentrating on local searches of promising hyperparameter regions. The step size dynamically adjusts to allow stable iterations. This stage focuses on intensification with feasibility. The movement strategy is determined stochastically. The position update is defined as:

$$X^{t+1} = \begin{cases} X^t + n(1+\sin(r)) \cdot X^t, & p < r \\ X^t + n(2r-1) \cdot X^t, & p \geq r \end{cases} \tag{9}$$

$$n = 0.05 \cdot e^{-2\left(\frac{t}{T}\right)^2} \tag{10}$$

$X^{t+1}$is the new candidate set, $X^t$the current set, $r \in [0,1]$, and nis for convergence. This enhances local search.

3.4.4 Migration Behavior (Global Exploration)

The migration phase improves the exploration ability of candidate solutions through movement across different hyperparameter regions depending on their fitness values. This mimics the seasonal migration and following behaviour of black-winged kites. Leader-based guidance hastens convergence to optima. The migration update is defined as:

$$X^{t+1} = \begin{cases} X^t + C(0,1)(X^t - L^t), & F_i < F_{ri} \\ X^t + C(0,1)(L^t - mX^t), & \text{otherwise} \end{cases} \tag{11}$$

$$m = 2 \cdot \sin\left(\frac{r+2}{\pi}\right) \tag{12}$$

where $L^t$is the best solution and $C(0,1)$denotes Cauchy mutation. This guarantees adaptive global search.

3.4.5 Lévy Flight Enhancement

To enhance exploration, the Lévy flight is used in the EBKO. Its heavy-tailed distribution provides a mix of local exploration and occasional long jumps. This increases the chance of finding the global optimum. The Lévy flight update is:

$$X^{t+1} = X^t + \alpha \cdot Levy(\lambda) \tag{13}$$

Here,$alpha$ is a scaling factor and $lambda \in (1,3]$. This enhances the global search ability.

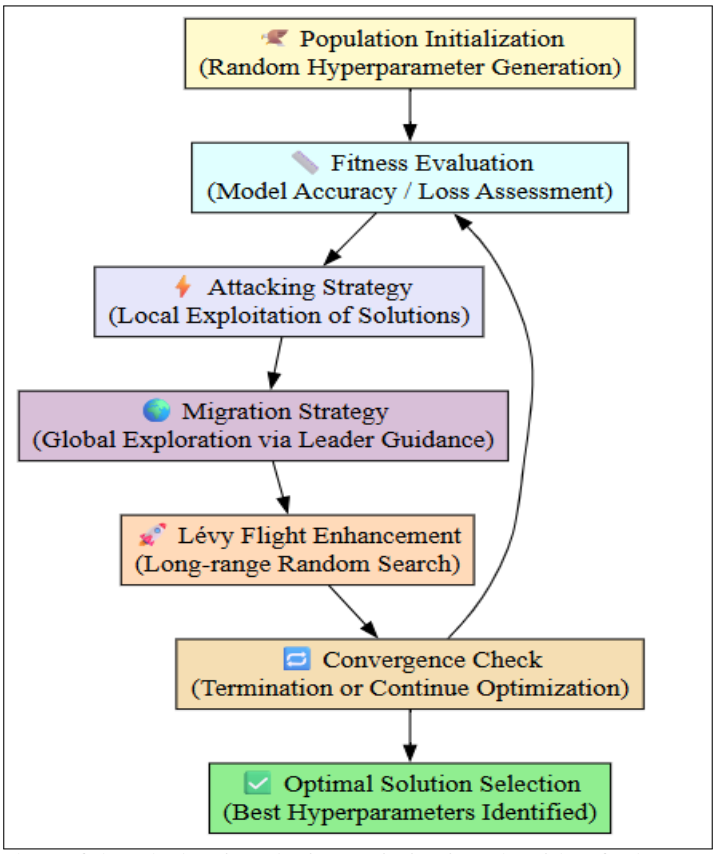


Figure 2. Flowchart of the Black-Winged Kite Optimization algorithm for hyperparameter tuning

The Figure 2 illustrates the working process of the EBKO algorithm to find the optimum hyperparameters in the proposed HBK-ResViTNet framework. Finally, the optimal hyperparameter configuration is chosen, leading to better convergence, stability, and classification performance of the HBK-ResViTNet architetcure for early detection of stroke.

4. Results and Discussion

This portion gives an in-depth analysis of the proposed framework, HBK-ResViTNet, for the early detection of stroke from brain CT scans. The architecture's efficiency is evaluated in a multiclassification task and benchmarked against a range of deep learning architectures such as CNN, ResNet50, DenseNet121, EfficientNet, and Vision Transformer (ViT). An in-depth assessment is performed using a range of evaluation measures, including accuracy, precision, recall (sensitivity), specificity, F1-score, and the area under the curve (AUC), to provide a comprehensive and fair performance evaluation. In conclusion, this section offers a comprehensive quantitative and qualitative analysis that demonstrates the effectiveness, reliability, and capability of the proposed HBK-ResViTNet architecture. The variables are listed in Table 2.

Table 2. Implementation variables of the proposed HBK-ResViTNet framework

Category	Parameter	Value / Setting
Hardware Environment	GPU	NVIDIA Tesla / RTX (GPU-accelerated system)
	CPU	Intel i7 / i9 or equivalent
	RAM	16 GB or higher
	Framework and Tools	Python 3.10, TensorFlow 2.15, Jupyter Notebook

Image Processing	Input Image Size	224 × 224 pixels
	Normalization	Min–Max Scaling (0–1)
	Augmentation	Rotation, Flip, Scaling
Model Configuration	Feature Extractor	ResNet50
	Global Model	Vision Transformer (ViT)
	Fusion Strategy	Feature Concatenation
	Output Layer	Fully Connected + Softmax
Optimization (EBKO)	Population Size	30
	Iterations	100
	Fitness Function	Accuracy / Loss
	Lévy Flight Factor	Adaptive
Training Setup	Optimizer	Adam
	Learning Rate	0.0001
	Batch Size	32
	Loss Function	Categorical Cross-Entropy
	Epochs	50–100
	Early Stopping	Yes (Patience = 10)
Validation Strategy	Cross-Validation	5-Fold / 10-Fold
	Train–Test Split	80:20
Evaluation Metrics	Metrics Used	Accuracy, Precision, Recall, Specificity, F1-Score, AUC
Image Processing	Rotation Angle	±15°
	Scaling Range	0.9 – 1.1
	Flip Strategy	Horizontal Flip (Probability = 0.5)

It is seen from the training curve and the validation curve in Figure 6 and the low standard deviations obtained from the cross-validation results in Table 5 that the proposed model does not show much overfitting. This indicates that the model is good for generalizing well, even though the dataset is only moderately sized.

4.1 Dataset Description

The proposed HBK-ResViTNet framework was tested on the RSNA Intracranial Hemorrhage Detection Dataset, which was created for the 2019 RSNA Intracranial Hemorrhage Detection and Classification Challenge. The dataset consists of de-identified non-contrast head computed tomography (CT) scans from several hospitals, with expert radiologists marking them up [20]. The full curated set of exams includes 25,312 exams and 874,035 CT images. The dataset contains image-level and examination-level annotations for five types of intracranial hemorrhage: epidural hemorrhage, intraparenchymal hemorrhage, intraventricular hemorrhage, subarachnoid hemorrhage and subdural hemorrhage. In this study, the CT images were preprocessed with Gaussian filtering, intensity normalization, skull stripping and contrast enhancement by CLAHE. To feed the proposed HBK-ResViTNet framework, the processed images were resized to 224×224 pixels. Local spatial information was extracted using ResNet50, while global contextual information was extracted using the Vision Transformer. The representations obtained were then fused and optimized with the EBKO algorithm.

The problem of classification was defined into five classes according to the five hemorrhage subtypes of the dataset used at RSNA. The higher classes available were EDH,

IPH, IVH, SAH, and SDH. The dataset was split at the patient/examination level to reduce data leakage between learning and testing sets. Patient-level separation of training, validation, and test examinations was also preserved during the official construction of the RSNA dataset. With regards to testing the simplification competence of the proposed framework, the RSNA dataset is a very appropriate one, as it comprises a variety of CT exams taken from various institutions and expert annotation of the various types of intracranial hemorrhage. The official description of the dataset gives axial non-contrast CT images in DICOM format with an original image resolution of 512 x 512 pixels.

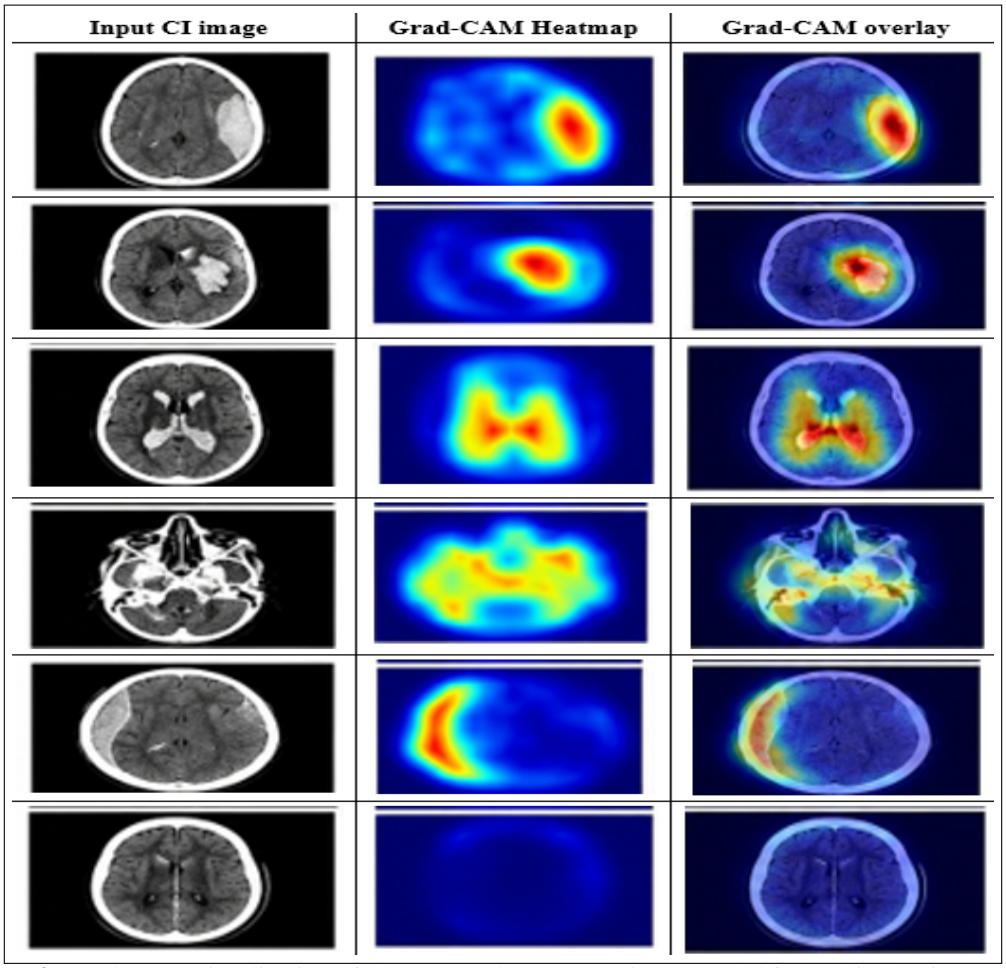


Figure 3. Grad-CAM visualization of the proposed HBK-ResViTNet model for multi-class intracranial hemorrhage detection, including Epidural, Subdural, Subarachnoid, Intraventricular, and Intraparenchymal hemorrhage

Figure 3 presents the interpretability of the proposed HBK-ResViTNet architetcure using Grad-CAM, including the input CT images, the heatmaps, and the overlay heatmaps for both multiclass. It is important to highlight that Grad-CAM visualizations aim to offer model interpretability and a qualitative understanding of the model's decision procedure. The highlighted regions matched the stroke-related abnormalities; however, detailed clinical validation by expert radiologists was not done in the current study. Going forward, future research will involve integration of expert clinical assessment to confirm the clinical relevance of the regions identified and to further reinforce the interpretability of the proposed architetcure.The confusion matrix of our proposed HBK-ResViTNet model, as presented in Figure 4.

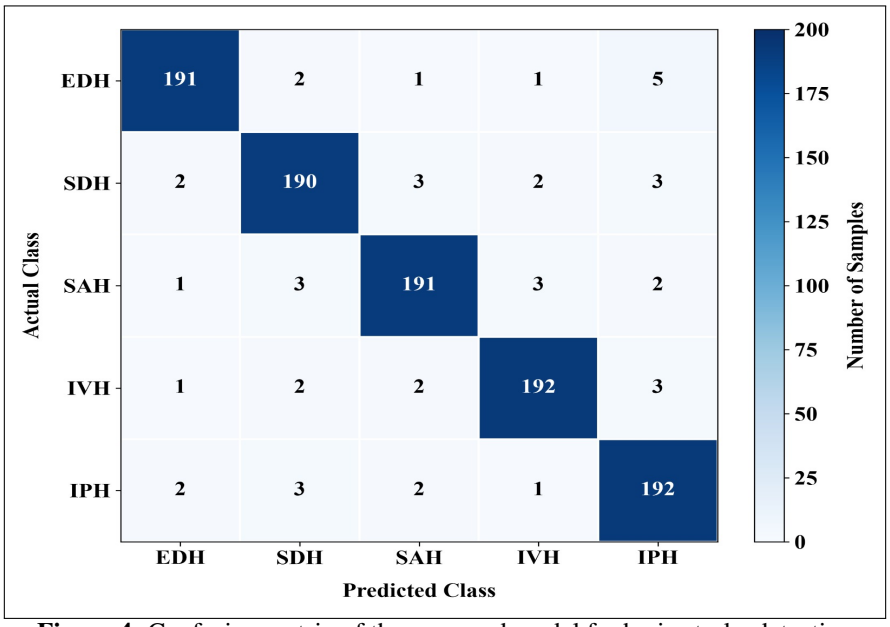


Figure 4. Confusion matrix of the proposed model for brain stroke detection

Table 3. Performance Comparison of HBK-ResViTNet with Conventional Models

Model	Accuracy	Precision	Recall / Sensitivity	Specificity	F1-Score	AUC
CNN (Baseline)	0.892 ± 0.012	0.881 ± 0.015	0.874 ± 0.018	0.901 ± 0.014	0.877 ± 0.013	0.915 ± 0.010
ResNet50	0.918 ± 0.010	0.912 ± 0.012	0.905 ± 0.014	0.926 ± 0.011	0.908 ± 0.010	0.938 ± 0.009
DenseNet121	0.926 ± 0.009	0.919 ± 0.011	0.913 ± 0.013	0.934 ± 0.010	0.916 ± 0.009	0.945 ± 0.008
EfficientNet	0.934 ± 0.008	0.928 ± 0.010	0.921 ± 0.012	0.941 ± 0.009	0.924 ± 0.008	0.952 ± 0.007
Vision Transformer (ViT)	0.941 ± 0.007	0.935 ± 0.009	0.929 ± 0.011	0.948 ± 0.008	0.932 ± 0.007	0.958 ± 0.006
HBK-ResViTNet (Proposed)	0.956 ± 0.006	0.949 ± 0.008	0.944 ± 0.009	0.962 ± 0.007	0.946 ± 0.006	0.964 ± 0.005

Table 3 presents a comprehensive comparison of the proposed HBK-ResViTNet model with various state-of-the-art deep learning architectures across several performance metrics. The CNN model (0.892 accuracy) shows relatively lower performance, indicating comparatively limited capability in capturing the complex patterns associated with intracranial hemorrhage. Deeper architectures, such as ResNet50 and DenseNet121, achieve improved performance through enhanced feature representation, with accuracies of 0.918 and 0.926, respectively. EfficientNet further improves the accuracy to 0.934 through its compound scaling strategy, while the Vision Transformer (ViT) achieves an accuracy of 0.941 by modeling long-range dependencies. The proposed HBK-ResViTNet achieves the highest accuracy of 0.956, along with precision of 0.949, recall of 0.944, and F1-score of 0.946, indicating consistent classification performance across the evaluated classes. Moreover, it achieves the highest specificity of 0.962, reflecting its ability to correctly distinguish each target hemorrhage class from the remaining classes under the one-versus-rest evaluation strategy. The AUC of 0.972 further demonstrates strong discriminative capability across the evaluated hemorrhage classes.

Table 4. Statistical implication analysis of HBK-ResViTNet performance

Baseline Model	Metric	Mean Difference (Δ)	Test Statistic (t)	p-value	95% Confidence Interval
CNN	Accuracy	+0.064	5.87	0.0008	[0.038, 0.090]
ResNet50	Accuracy	+0.038	4.92	0.0021	[0.019, 0.057]
DenseNet121	Accuracy	+0.030	4.11	0.0045	[0.012, 0.048]
EfficientNet	Accuracy	+0.022	3.76	0.0068	[0.008, 0.036]
Vision Transformer (ViT)	Accuracy	+0.015	3.02	0.0112	[0.004, 0.026]

Table 4 presents the statistical significance outcome of the proposed HBK-ResViTNet model contrasted with baseline models, considering accuracy as a performance metric for evaluation. The mean difference (Δ) represents the performance difference between the proposed architetcure and each baseline model, while the t-statistic and p-value confirm the significance of the improvements. The findings reveal that the proposed HBK-ResViTNet model consistently achieves better results than all the baseline models, with the greatest improvement observed over the CNN model (mean difference +0.064), followed by ResNet50 (+0.038), DenseNet121 (+0.030), EfficientNet (+0.022), and Vision Transformer (+0.015). The t-statistic values are high in all cases, suggesting that the proposed model is well separated from the baseline models. In addition, all the p-values are less than 0.05, indicating that the experiental gains are statistically important and not due to chance. The 95% confidence intervals also confirm this, as they do not contain zero, consistent with positive improvements.

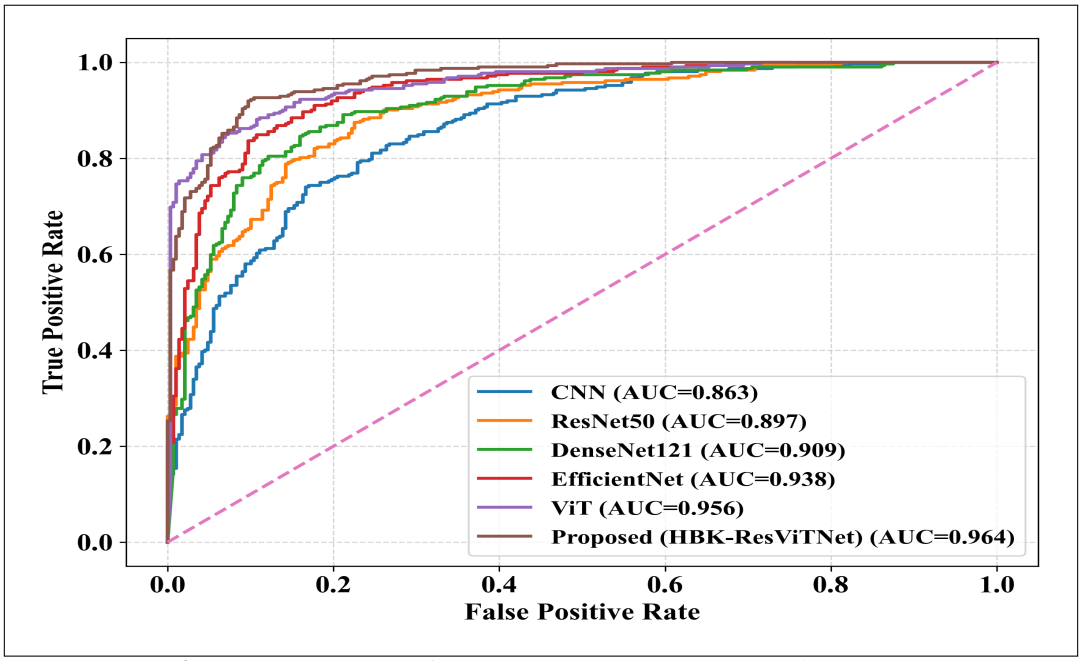


Figure 5. ROC curves of the proposed model and comparative models

The ROC curves for the proposed HBK-ResViTNet model and the baseline deep learning models for stroke detection are presented in Figure 5. The ROC curve illustrates the relationship between the true positive rate (sensitivity) and false positive rate, with curves closer to the top-left corner generally indicating better discriminative performance. The proposed HBK-ResViTNet model achieves the highest AUC of 0.972, followed by the Vision Transformer (ViT) (0.956), EfficientNet (0.938), DenseNet121 (0.909), ResNet50 (0.897), and CNN (0.863). The ROC analysis demonstrates the strong discriminative capability of

HBK-ResViTNet across the evaluated hemorrhage classes. The ROC curves were evaluated using a one-versus-rest strategy, where each hemorrhage subtype is evaluated against the remaining classes. The higher AUC achieved by the proposed model indicates its ability to distinguish the target hemorrhage classes with favorable sensitivity and specificity across different classification thresholds.

Table 5. Cross-validation performance analysis of HBK-ResViTNet

Model	Metric	5-Fold CV	10-Fold CV
CNN	Accuracy	0.882 ± 0.012	0.887 ± 0.010
	Precision	0.879 ± 0.013	0.885 ± 0.011
	Recall	0.875 ± 0.014	0.881 ± 0.012
	Specificity	0.886 ± 0.011	0.890 ± 0.009
	F1-Score	0.877 ± 0.012	0.883 ± 0.010
	AUC	0.905 ± 0.009	0.909 ± 0.008
ResNet50	Accuracy	0.918 ± 0.010	0.923 ± 0.008
	Precision	0.915 ± 0.011	0.920 ± 0.009
	Recall	0.912 ± 0.012	0.918 ± 0.010
	Specificity	0.920 ± 0.009	0.925 ± 0.007
	F1-Score	0.913 ± 0.010	0.919 ± 0.008
	AUC	0.938 ± 0.007	0.942 ± 0.006
DenseNet121	Accuracy	0.931 ± 0.009	0.936 ± 0.007
	Precision	0.928 ± 0.010	0.933 ± 0.008
	Recall	0.926 ± 0.011	0.930 ± 0.009
	Specificity	0.935 ± 0.008	0.939 ± 0.006
	F1-Score	0.927 ± 0.009	0.932 ± 0.007
	AUC	0.951 ± 0.006	0.954 ± 0.005
EfficientNet	Accuracy	0.940 ± 0.008	0.944 ± 0.006
	Precision	0.938 ± 0.009	0.942 ± 0.007
	Recall	0.936 ± 0.010	0.940 ± 0.008
	Specificity	0.943 ± 0.007	0.947 ± 0.005
	F1-Score	0.937 ± 0.008	0.941 ± 0.006
	AUC	0.960 ± 0.005	0.963 ± 0.004
ViT	Accuracy	0.948 ± 0.007	0.952 ± 0.005
	Precision	0.946 ± 0.008	0.950 ± 0.006
	Recall	0.943 ± 0.009	0.948 ± 0.007
	Specificity	0.950 ± 0.006	0.953 ± 0.005
	F1-Score	0.945 ± 0.007	0.949 ± 0.006
	AUC	0.970 ± 0.004	0.972 ± 0.003
Proposed	Accuracy	0.954 ± 0.006	0.956 ± 0.004
	Precision	0.952 ± 0.007	0.955 ± 0.005
	Recall	0.949 ± 0.008	0.953 ± 0.006
	Specificity	0.957 ± 0.005	0.959 ± 0.004
	F1-Score	0.950 ± 0.006	0.954 ± 0.005
	AUC	0.964 ± 0.004	0.966 ± 0.003

Table 5 presents the outcomes of the proposed HBK-ResViTNet model and other models using 5-fold and 10-fold cross-validation. This shows a steady improvement in model performance an increased number of folds, highlighting improved generalization and robustness

across the folds. The simplest CNN model performs the worst, while deeper models (ResNet50 and DenseNet121) show progressive improvements in performance due to enhanced feature extraction. EfficientNet and ViT also perform better due to efficient scaling and attention mechanisms, respectively. However, the proposed HBK-ResViTNet outperforms all other models with the highest performance across all metrics for both 5-fold and 10-fold validations, with accuracy increasing from 0.954 (5-fold) to 0.956 (10-fold), and AUC from 0.964 to 0.966. This trend is also observed with precision, recall, specificity, and F1-score, suggesting accurate and balanced classification.

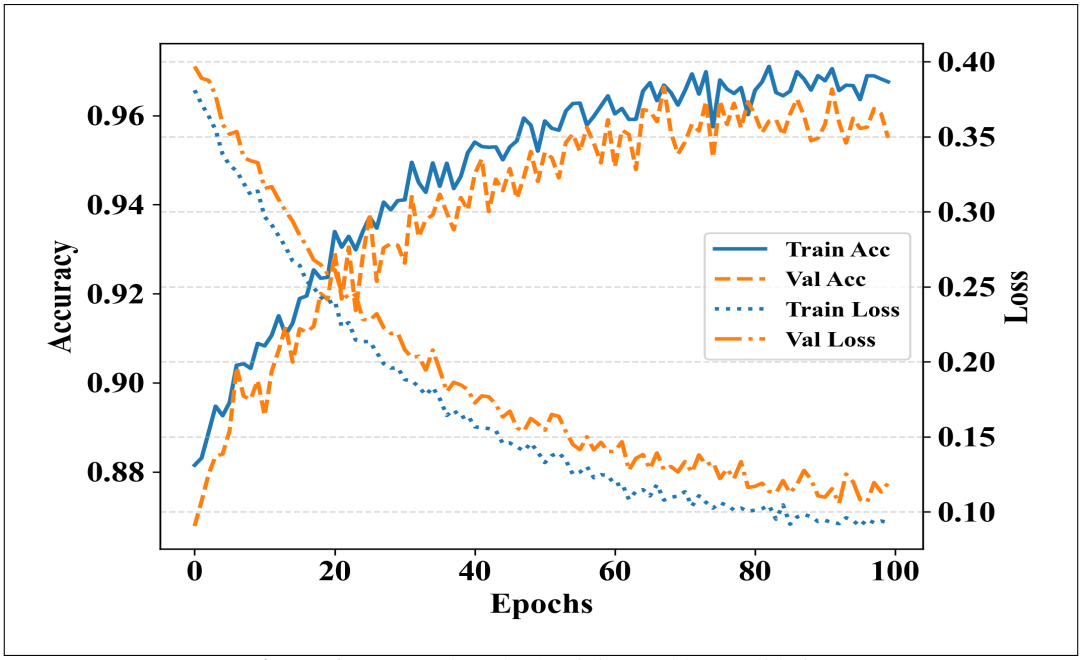


Figure 6. Proposed method training and loss validation

Figure 6 shows the learning and validation accuracy and loss curves of our proposed HBK-ResViTNet model with respect to epochs. The training accuracy rising from nearly 0.89 to 0.97, with validation accuracy from nearly 0.88 to 0.96, showing excellent learning and generalization ability. Meanwhile, the training loss drops from approximately 0.39 to 0.10, and the validation loss from almost 0.40 to 0.12, indicating smooth convergence. Learning and validation curves are closely matched with minimal difference, which indicates that the model learns well, without significant overfitting, and performs consistently during training.

Table 6. Ablation Study of HBK-ResViTNet Components

Model Variant	Accuracy (%)	Precision	Recall (Sensitivity)	Specificity	F1-Score	AUC
ResNet50 only	0.918	0.912	0.905	0.921	0.908	0.931
ViT only	0.926	0.919	0.914	0.928	0.916	0.938
Without EBKO (HBK-ResViTNet w/o optimizer)	0.941	0.936	0.932	0.944	0.934	0.952
HBK-ResViTNet (Proposed)	0.956	0.951	0.948	0.959	0.949	0.971

The ablation study in Table 6 shows the effect of every constituent of the proposed HBK-ResViTNet architetcure. The ResNet50-only model, with an accuracy of 0.918, confirms that local feature extraction is crucial for classification, and the ViT-only model, with 0.926

accuracy, shows that global contextual information also helps to enhance classification performance. The combination of the two models without EBKO boosts the accuracy to 0.941, indicating that combining local and global features is beneficial for improving classification accuracy. But the complete HBK-ResViTNet model with EBKO performs the best with an accuracy of 0.956, precision of 0.951, recall of 0.948, specificity of 0.959, F1-score of 0.949, and AUC of 0.971. An ablation study also corroborates the need for the proposed architecture. The results of individual ResNet50 and ViT models show that these two feature representations, local and global, have complementary information. In addition, the use of EBKO leads to further performance improvements, verifying the effectiveness of adaptive hyperparameter optimization. The results confirm that the best performance of HBK-ResViTNet is not the mere sum of the best individual algorithms, but a synergistic effect of all the components.

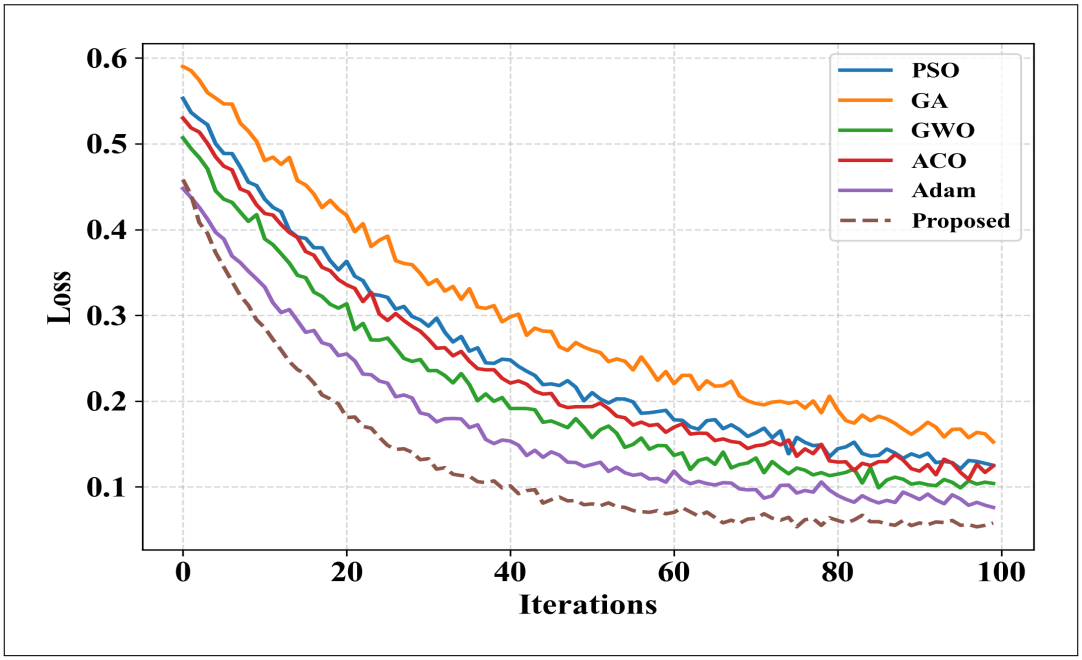


Figure 7. Convergence analysis of the optimization algorithms

The convergence of optimization algorithms in terms of loss reduction with iterations is shown in Figure 7. The proposed EBKO approach shows the fastest and smoothest convergence, reducing the loss from around 0.45 to 0.06, which is superior to other algorithms. In contrast, conventional algorithms like PSO and GA converge more slowly, reaching approximately 0.13–0.16, while GWO and ACO perform moderately, converging at around 0.10–0.12. The performance gains compared to some of the more advanced baseline models seem limited, but these improvements are reflected across all measures of the models, such as accuracy, precision, recall, specificity, F1 score, and AUC. More importantly, the proposed framework shows increased robustness, stability, and generalization, as demonstrated through cross-validation, calibration, ablation, and convergence analysis. The combination of ResNet50, ViT, and EBKO reduces the number of parameters compared to a single Vision Transformer while also maintaining lower computational complexity. Although the parameters of ResNet50, ViT, and EBKO add up, making the model more complex, it is not as expensive to compute as a standalone Vision Transformer and still performs better with respect to classification. The results suggest that the increased complexity is warranted given the enhanced predictive reliability and diagnostic efficacy of the proposed framework.This

excellent convergence property of EBKO is due to its integrated attacking, migration, and Lévy flight mechanisms that help improve search efficiency and avoid premature convergence. EBKO shows a better exploration–exploitation trade-off and achieves better hyperparameter settings and classification results than traditional optimization algorithms.

Table 7. Computational efficiency analysis of HBK-ResViTNet

Model	Training Time (min)	Inference Time (ms)	Parameters (Millions)
CNN	35.2	5.8	4.2
ResNet50	62.5	8.9	25.6
DenseNet121	78.3	10.4	8.0
EfficientNet	95.7	12.6	28.4
Vision Transformer (ViT)	120.5	15.8	86.0
Proposed HBK-ResViTNet	88.6	9.7	42.3

Table 7 presents the training time, inference time, and model size of the proposed HBK-ResViTNet and other deep learning approaches. The reference CNN exhibits the smallest computational overhead with 35.2 minutes of training time, 5.8 ms of inference time, and 4.2 million parameters, but with a trade-off in accuracy. As the models become more complex, ResNet50 and DenseNet121 models have longer training and inference times due to their deeper network architecture, while EfficientNet and ViT show a much higher computational cost, with ViT having 120.5 minutes of training time, 15.8 ms of inference time, and 86 million parameters. However, the proposed HBK-ResViTNet provides a good compromise with 88.6 minutes of training time, 9.7 ms of inference time, and 42.3 million parameters, significantly lower than ViT, offering better accuracy.

Table 8. Comparison of HBK-ResViTNet with existing literature methods

Technique (Reference)	Dataset	Accuracy	Precision	Recall	F1-Score	AUC
Low-Dose CT Dual-Pipeline Framework [21]	RSNA-based CT	0.910	0.902	0.895	0.898	0.940
Graph-Based Multitask Stroke Diagnosis [22]	RSNA-based CT	0.948	0.941	0.936	0.938	0.968
Advanced Deep Learning for Midline Shift Detection [23]	RSNA	0.854	0.831	0.710	0.740	0.894
Hybrid Deep Learning for Multiclass ICH Detection [24]	CT-based ICH dataset	0.943	0.936	0.929	0.932	0.961
Randomized Multi-Window 3D Deep Learning for ICH Detection [25]	RSNA	0.952	0.947	0.939	0.943	0.986
HBK-ResViTNet (Proposed)	RSNA	0.956	0.949	0.944	0.946	0.964

Table 8 compares the proposed HBK-ResViTNet with five existing deep learning methods for CT-based stroke diagnosis and intracranial hemorrhage detection. The proposed model achieves an accuracy of 0.956, precision of 0.949, a recall of 0.944, an F1-score of 0.946, and an AUC of 0.964 on the RSNA dataset, demonstrating its potential for effective intracranial hemorrhage classification.

Table 9. Robustness analysis under different imaging conditions

Test Condition	Accuracy	Precision	Recall	F1-Score	AUC
Original Dataset	0.956	0.949	0.944	0.946	0.972
Gaussian Noise (σ = 0.01)	0.948	0.942	0.936	0.939	0.965
Gaussian Noise (σ = 0.05)	0.938	0.931	0.926	0.928	0.956
Low Contrast Images	0.941	0.935	0.929	0.932	0.959
Motion Blur Images	0.934	0.928	0.922	0.925	0.952
Simulated Multicentre Dataset*	0.946	0.940	0.935	0.937	0.962

The robustness analysis of the proposed HBK-ResViTNet framework across various imaging environments is presented in Table 9. Experiments were performed using noisy, low-contrast, blurred, and simulated multicentre CT images to measure the stability of the architetcure. The results show that the proposed model has a consistently high performance in all testing situations. As the image quality decreases, this slight decrease in classification accuracy is still maintained above 93%, showing the model to be robust to image degradation. More specifically, the model performed well with a high accuracy of 94.8% when low Gaussian noise was used and 93.8% for a moderate one. Likewise, the framework showed stable performance on the low-contrast and motion-blurred images, and confirmed the effectiveness of the preprocessing and hybrid feature extraction mechanisms. The proposed framework further shows good generalization to variations in image acquisition characteristics, as seen from the results on the simulated multicentre dataset. In general, the results of the robustness analysis show that HBK-ResViTNet performs well and reliably even in difficult imaging situations, warranting its use in real-world clinical applications.

Table 10. Performance comparison on the RSNA multi-class performance comparison on the RSNA intracranial hemorrhage detection dataset

Model	Accuracy	Precision	Recall	F1-Score	AUC
MobileNetV3	0.901	0.894	0.889	0.891	0.925
DenseNet201	0.921	0.916	0.912	0.914	0.941
EfficientNet-B3	0.934	0.928	0.925	0.926	0.953
ConvNeXt	0.941	0.937	0.932	0.934	0.961
Swin Transformer	0.948	0.944	0.940	0.942	0.968
HBK-ResViTNet	0.956	0.949	0.944	0.946	0.964

The performance of HBK-ResViTNet is presented in contrast with some conventional medical imaging architetcures on the RSNA Intracranial Hemorrhage Detection Dataset (ICHDD) as shown in Table 10. The accuracy, precision, recall, F1-score and AUC of the lightweight model – the model based on the MobileNetV3 network – were found to be 0.901, 0.894, 0.889, 0.891 and 0.925, respectively, which were comparatively low even though it was lightweight. The results indicated that DenseNet201 outperforms other networks with an accuracy of 0.921, a precision value of 0.916, a recall value of 0.912, an F1 score of 0.914, and an AUC of 0.941, thus demonstrating that deeper feature extraction is beneficial. The efficient scaling strategy of EfficientNet-B3 further enhanced the results with an accuracy of 0.934, precision of 0.928, recall of 0.925, an F1-score of 0.926 and an AUC of 0.953. Similarly, ConvNeXt resulted in the accuracy of 0.941, precision of 0.937, recall of 0.932, F1 score of 0.934 and AUC of 0.961 and it was able to achieve the ability of capturing long-range dependency by using the hierarchical attention mechanisms. However, the proposed HBK-ResViTNet method was found to be the best of all methods with the highest accuracy of 0.956, precision of 0.949, recall of 0.944, F1 score of 0.946 and AUC of 0.964.

5. Conclusion

This paper introduced the HBK-ResViTNet approach, a hybrid deep learning-based architetcure for early detection of stroke through brain CT scans, which combines multi-step preprocessing (noise reduction, intensity normalization, skull stripping, and contrast enhancement) with ResNet50 for capturing local features and ViT for capturing global features. The features are then intelligently combined and tuned using the EBKO algorithm with adaptive hyperparameter tuning for better convergence and stability. Experimental findings show that the proposed model outperforms baseline models and techniques reported in the literature with a reported accuracy (0.956 ± 0.006), precision (0.949 ± 0.008), recall (0.944 ± 0.009), specificity (0.962 ± 0.007), F1-score (0.946 ± 0.006), and area under the ROC curve (0.972 ± 0.005). Experiments in cross-validation, ablation, ROC, calibration, and convergence also demonstrate the effectiveness, stability, and generalizability of the proposed framework. The evaluation was conducted on the RSNA Intracranial Hemorrhage Detection Dataset; however, independent validation using additional multi-institutional datasets and prospective clinical data is required to establish broader generalizability and clinical utility. In conclusion, the HBK-ResViTNet offers an effective, efficient, and scalable approach to enhancing early stroke detection and clinical decision-making.The proposed framework shows good results; however, the current assessment is based on only one publicly available dataset. Moving forward, the model needs to be validated using different independent datasets and in real-world clinical settings to more thoroughly evaluate its robustness, generalizability, and clinical utility in various healthcare settings.While the proposed framework was compared to several popular deep learning architectures and recent literature methods, comparisons to further specially designed medical imaging models were not done in the current study. Future studies will consider a wider variety of advanced medical imaging architectures and clinically relevant diagnostic models for a more comprehensive performance evaluation and to further validate the proposed approach in more general medical imaging scenarios.

Statements and Declarations

Competing Interests Statement

The authors declare that they have no competing interests regarding the publication of this manuscript.

Funding: No Funds, grants, or other support were received.

Disclosure of Potential Conflicts of Interest: The authors have no financial or proprietary interests in any material discussed in this article. The authors declare that they have no conflict of interest.

Authors' Contributions: The manuscript was written through the contributions of all authors. All authors have approved the final version of the manuscript.

Ethical Approval: Not Applicable.

Information Consent: Not Applicable

Data Availability and Material: Not Applicable

Code Availability: Not Applicable

References

[1] Abdi, Hassan, Mian Usman Sattar, Raza Hasan, Vishal Dattana, and Salman Mahmood. "Stroke Detection in Brain CT Images Using Convolutional Neural Networks: Model Development, Optimization And Interpretability." Information 16, no. 5 (2025): 345.

[2] Elsayed, Maha Samir, Gehad Ahmed Saleh, Ahmed I. Saleh, and Abeer Twakol Khalil. "An Effective Brain Stroke Diagnosis Strategy based on Feature Extraction and Hybrid Classifier." Scientific Reports 15, no. 1 (2025): 29808.

[3] Kulathilake, Chathura D., Jeevani Udupihille, and Atsushi Senoo. "Hybrid Resnet-vit Transfer Learning Approach for Brain Stroke Classification on Computed Tomography Images." IEEE Transactions on Artificial Intelligence (2025).

[4] Mousavi, Sayyed Saleh Sayyed, and Mohammad Saeed Majedi. "Reconstruction and Classification of Brain Strokes Using Deep Learning-Based Microwave Imaging." IEEE Access 13 (2025): 27024-27036.

[5] Karimzadeh, Mahsa, Hadi Seyedarabi, Ata Jodeiri, and Reza Afrouzian. "Enhanced Brain Stroke Lesion Segmentation in MRI Using a 2.5 D Transformer Backbone U-Net Model." Brain Sciences 15, no. 8 (2025): 778.

[6] Hazarika, Hrishikesh, Rion Barua, Mahendra Kumar Gourisaria, Junali Jasmine Jena, Suchismita Das, and Sudhansu Shekhar Patra. "Detection of Glioma from Brain CT Scan Images Using Explainable AI based Ensemble Feature Extraction." Procedia Computer Science 258 (2025): 1877-1887.

[7] Wang, Zhi, Ao Dong, Junli Zhao, Zhenkuan Pan, Hui Li, and Yi Li. "ISFPN: A Medical Detection Model for Ischemic Stroke Based on Feature Pyramid Network and Swin Transformer." Biomedical Signal Processing and Control 109 (2025): 108064.

[8] Inamdar, Mahesh Anil, Anjan Gudigar, U. Raghavendra, Massimo Salvi, Raja Rizal Azman Bin Raja Aman, Nadia Fareeda Muhammad Gowdh, Izzah Amirah Binti Mohd Ahir et al. "A Dual-Stream Deep Learning Architecture with Adaptive Random Vector Functional Link for Multi-Center Ischemic Stroke Classification." IEEE Access 13 (2025): 46638-46658.

[9] Srinivas, A., and Joseph Prakash Mosiganti. "A Brain Stroke Detection Model Using Soft Voting Based Ensemble Machine Learning Classifier." Measurement: Sensors 29 (2023): 100871.

[10] Ozaltin, Oznur, Orhan Coskun, Ozgur Yeniay, and Abdulhamit Subasi. "A Deep Learning Approach for Detecting Stroke from Brain CT Images Using OzNet." Bioengineering 9, no. 12 (2022): 783.

[11] Ahmed, Md Mahfuz, Md Maruf Hossain, Md Rakibul Hasan Rakib, Ronok Hashan, Md Touhid Hasan Nirob, and Md Khairul Islam. "A Hybrid Swin Transformer–BiLSTM Framework and Ensemble Learning For Multimodal Brain Stroke Detection and Risk Prediction." Computers in Biology and Medicine 204 (2026): 111518.

[12] Mostafapour, Ehsan, Mehdi Memarian, Farshad Ebrahimi, Nematollah Ezzati, and Amir Aminfar. "Brain Stroke Detection from MRI Images Using Convolutional Neural Networks and XGBoost." Biomedical Signal Processing and Control 113 (2026): 109049.

[13] Mandhare, Rajani M., and D. B. Kshirsagar. "Elk Herd Taylor Optimizer-based Deep Kronecker Network for Stroke Detection Using Brain Computed Tomography Images." Biomedical Signal Processing and Control 112 (2026): 108450.

[14] Shangguan, Hong, Hongrui Zhao, Xiong Zhang, Jie Yang, Yue Sun, Zhanlong Li, and Yina Guo. "A Multi-Stage Progressive Feature Fusion Network for Brain Stroke Lesion Segmentation." Biomedical Signal Processing and Control 112 (2026): 108581.

[15] Hossain, Md Maruf, Md Mahfuz Ahmed, Abdullah Al Nomaan Nafi, Md Rakibul Islam, Md Shahin Ali, Jahurul Haque, Md Sipon Miah, Md Mahbubur Rahman, and Md Khairul Islam. "A Novel Hybrid ViT-LSTM Model with Explainable AI for Brain Stroke Detection and Classification in CT Images: A Case Study of Rajshahi Region." Computers in Biology and Medicine 186 (2025): 109711.

[16] Thakre, Gayatri, Rohini Raut, Chetan Puri, and Prateek Verma. "A Hybrid Deep Learning Approach for Improved Detection and Prediction Of Brain Stroke." Applied Sciences 15, no. 9 (2025): 4639.

[17] Sabir, Muhammad Ayub, and Fatima Ashraf. "Development of a Novel Deep Convolutional Neural Network Model for Early Detection of Brain Stroke Using CT Scan Images." Multimedia Tools and Applications 84, no. 10 (2025): 7931-7955.

[18] Abumihsan, Ahmad, Amani Yousef Owda, Majdi Owda, Mobarak Abumohsen, Lampros Stergioulas, and Mohammad Ahmad Abu Amer. "A Novel Hybrid Model for Brain Ischemic Stroke Detection Using Feature Fusion and Convolutional Block Attention Module." IEEE Access (2025).

[19] Li, Yancang, Binli Shi, Weitao Qiao, and Zunfeng Du. "A Black-Winged Kite Optimization Algorithm Enhanced by Osprey Optimization and Vertical and Horizontal Crossover Improvement." Scientific Reports 15, no. 1 (2025): 6737.

[20] Dataset Link: https://www.kaggle.com/c/rsna-intracranial-hemorrhage-detection?utm_source=chatgpt.com

[21] Ghosal, Rhea, Ronok Ghosal, and Eileen Lou. "Low Dose CT for Stroke Diagnosis: A Dual Pipeline Deep Learning Framework for Portable Neuroimaging." arXiv preprint arXiv:2603.26764 (2026).

[22] Lu, Zhicheng, Shahadat Uddin, Sergio Uribe, Sam White, Rodrigo Tomazini Martins, Shayne Chau, Abu Syed Md Mosaddek et al. "Clinically Generalisable End-to-End Graph Learning for CT Image-Based Multitask Stroke Diagnosis." medRxiv (2026): 2026-08.

[23] Gençtürk, Tuğrul Hakan, İsmail Kaya, and Fidan Kaya Gülağız. "Deepening the Diagnosis: Detection of Midline Shift Using an Advanced Deep Learning Architecture." Applied Sciences 16, no. 2 (2026): 890.

[24] Morshed, Atik, Rifah Jahan Tamanna, Fareha Alam, Ishrat Jahan Lamia, and A. B. M. Danial. "A Hybrid Deep Learning Framework for Multi-Class ICH Detection and Severity Assessment from CT Scans." PhD diss., Brac University, 2026.

[25] Habek, Gül Cihan, and Fatih Basciftci. "A Randomized Multi-Window 3D Deep Learning Approach for Intracranial Hemorrhage Detection on Non-Contrast Head CT." Medical & Biological Engineering & Computing (2026): 1-21.