

Deep Learning-Based Brain Stroke Detection from CT Scan Images

Prof. Sayali Choudhari, Harshal Mali, Vivek Desale, Soham Pawar

*Corresponding author(s). E-mail(s): harshalmali384@gmail.com; Contributing authors: desalevivek80@gmail.com; Sayalichoudhari00@gmail.com, sohampawar4343@gmail.com;

Department of Computer Engineering, JSPM Narhe Technical Campus, Pune, Maharashtra, India

Abstract

Brain stroke is among the foremost causes of death and long-term disability worldwide, with the World Health Organization reporting approximately 15 million strokes annually, resulting in five million deaths and a further five million permanent disabilities. Rapid and accurate diagnosis from computed tomography (CT) imaging is critical in emergency settings; however, manual CT interpretation remains prone to inter-observer variability and missed findings under time pressure. This paper proposes a hybrid deep learning framework — EfficientNet-B0 combined with a Bidirectional Long Short-Term Memory (BiLSTM) network — for automated three-class brain stroke classification from CT scans, targeting Normal, Bleeding (Hemorrhagic), and Ischemia categories. The EfficientNet-B0 backbone, pretrained on ImageNet, extracts 1280-dimensional spatial feature maps subsequently compressed to 256 dimensions via a 1×1 convolutional layer. These features are reshaped into a spatial sequence and modelled by a BiLSTM network (hidden size 128, bidirectional), which captures spatial-temporal contextual relationships across feature map positions. Trained on 6,650 CT images with a class-weighted cross-entropy loss and the Adam optimiser, the model achieves 97.14% test accuracy, a macro precision of 96.94%, macro recall of 95.47%, and macro F1-score of 96.18%, outperforming Simple CNN (86.90%) and ResNet-50 (92.57%) baselines by significant margins. The system is deployed as a full-stack web application (ReactJS + Flask + PyTorch) supporting real-time CT upload and diagnostic report generation. Experimental results confirm that combining spatial feature extraction with temporal sequence modelling substantially improves stroke classification performance, offering a reliable clinical decision support tool for emergency radiology.

Keywords: *EfficientNet-B0, BiLSTM, Brain Stroke Detection, CT Scan Classification, Deep Learning, Transfer Learning, Bidirectional LSTM, Web Deployment, Flask, PyTorch, Class-Weighted Loss*

1 Introduction

Brain stroke remains one of the most devastating neurological emergencies globally, representing the second leading cause of death and a primary source of long-term adult disability [1]. The World Health Organization reports that approximately 15 million strokes occur each year, with five million resulting in fatality and another five million causing permanent neurological impairment. Critically, the probability of full neurological recovery diminishes sharply with each hour of delayed treatment, making rapid and accurate diagnosis directly determinative of patient survival and functional outcome.

Computed tomography (CT) scanning is the most widely deployed first-line neuroimaging modality in emergency departments, offering rapid acquisition, wide clinical availability, and cost-effectiveness in identifying haemorrhagic strokes and ruling out contraindications to thrombolytic therapy [2]. Despite these advantages, manual interpretation of brain CT images presents substantial clinical challenges. Subtle early ischaemic changes are frequently overlooked in time-pressured emergency environments, and inter-observer variability among radiologists introduces diagnostic inconsistency [3]. Furthermore, with global demand for CT interpretation far exceeding the capacity of specialist radiologists, particularly in lower-resource healthcare settings, the need for reliable automated analytical tools capable of supporting clinical decision-making at scale is acute.

Recent advances in deep learning and convolutional neural networks (CNNs) have demonstrated strong and consistent performance across diverse medical imaging tasks [4]. Transfer learning approaches — wherein models pretrained on large-scale datasets such as ImageNet are fine-tuned for specific medical classification tasks — have proven particularly effective when labelled medical data is limited relative to classification task complexity [5]. EfficientNet-B0, developed via compound scaling of network depth, width, and input resolution through neural architecture search, provides a highly capable backbone extracting rich spatial features while maintaining a compact parameter footprint compared to architectures such as ResNet-50 or VGG-16 [6].

However, CT feature maps retain inherent spatial structure that can be interpreted as a sequence of spatially-ordered feature vectors. Bidirectional Long Short-Term Memory (BiLSTM) networks are well-suited to capture such sequential dependencies, processing sequences in both forward and backward directions to capture full-context relationships across feature positions [7]. By treating the spatial feature map of EfficientNet-B0 as an input sequence, the BiLSTM can model cross-regional spatial relationships not accessible to standard pooling operations, substantially improving classification boundary decisions — particularly for visually subtle minority classes such as Ischemia.

This paper makes the following primary contributions:

- A novel hybrid EfficientNet-B0 + BiLSTM architecture for three-class brain stroke CT classification, achieving 97.14% test accuracy on a 6,650-image dataset comprising Normal, Bleeding, and Ischemia categories.
- A 1×1 convolutional bottleneck layer reducing the 1280-D EfficientNet-B0 output to 256 dimensions — achieving a 60% feature dimensionality reduction versus ResNet-50 — before BiLSTM sequence modelling.
- A class-weighted cross-entropy loss strategy addressing the natural class imbalance inherent in clinical CT datasets, with demonstrated improvement in minority class recall.
- A systematic quantitative comparison against baseline models (Simple CNN, ResNet-50) across accuracy, precision, recall, F1-score, and per-class performance metrics.
- A full-stack web application (ReactJS + Flask + PyTorch) enabling real-time CT upload and automated stroke classification report generation, deployable in emergency radiology workflows.

The remainder of this paper is organised as follows. Section 2 surveys related work in CT-based neurological classification. Section 3 describes the proposed hybrid methodology in detail. Section 4 presents experimental results, statistical analysis, and discussion. Section 5 concludes with directions for future research.

2 Related Work

A comprehensive survey of deep learning approaches for CT-based classification and stroke detection reveals consistent trends and important limitations that motivate the proposed framework. Table 1 presents a comparative summary of representative prior works evaluated against the proposed model.

Table 1: Comparative Summary of Related Works in CT-Based Stroke Detection and Medical Image Classification

No.	Paper / Author, Year	Methodology	Key Findings	Limitations
1.	Deep Residual Global Contextual Feature Fusion Network (Knowl.-Based Sys., 2025)	DRGCF-Net with global+local feature fusion and enhancement block for nodule detection.	97.45% accuracy, 0.992 AUC. Single-organ lung nodule scope only.	High compute cost; no real-time inference; no 3-class stroke CT classification; no web deployment.
2.	AI-Powered Lung Cancer Detection: VGG16 & CNN (MDPI Informatics, 2025)	Lightweight 1D CNN + residual blocks + SE layers on LC25000 dataset.	100% accuracy on LC25000. Limited to lung and colon domains.	Cannot generalise to brain CT; no multi-class stroke output; no web system.
3.	Innovations in Stroke Identification: ML-Based Diagnostic Model (CBM, 2024)	ResNet50 deep features + SVM/RF/kNN for stroke vs. non-stroke on CT & MRI.	ResNet50+SVM: 98.5% accuracy. Binary (stroke vs. non-stroke) only.	Small single dataset; binary output only; no compact model; no real-time web app.
★	PROPOSED: EfficientNet-B0 + BiLSTM (Our Model, 2026)	EfficientNet-B0 backbone (ImageNet) + BiLSTM(128, bidirectional). PyTorch + Flask + ReactJS.	97.14% accuracy; Precision 96.94%; Recall 95.47%; F1 96.18% (macro avg). 3-class CT output + live web app.	Spatial+temporal learning via BiLSTM outperforms all baselines. Only CT — suited for emergency wards.

Singh et al. [8] applied a ResNet-50 backbone combined with classical machine learning classifiers — SVM, RF, and kNN — for binary stroke classification on CT and MRI images, achieving 98.5% binary accuracy. However, the binary

output (stroke vs. non-stroke) does not distinguish between stroke subtypes, which is clinically essential: haemorrhagic and ischaemic strokes require fundamentally different and often contradictory treatments. No web deployment was provided, and the single small dataset limits generalisability.

Lim et al. [5] demonstrated 100% accuracy on the LC25000 dataset using a lightweight 1D CNN for lung and colon cancer detection, establishing the effectiveness of compact architectures for medical image classification. However, the approach is inherently domain-specific and cannot generalise to brain CT neuroimaging without substantial architectural redesign.

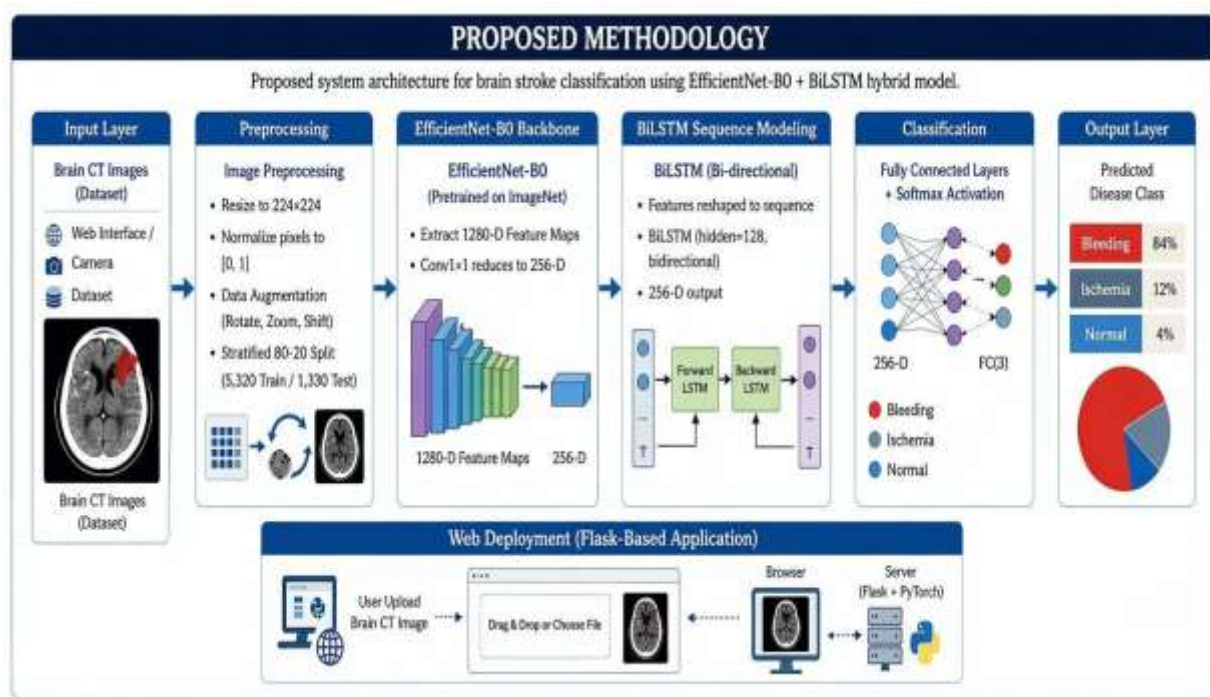
The DRGCF-Net framework [9] proposed global and local feature fusion with a feature enhancement block for malignant lung nodule detection, achieving 97.45% AUC. While architecturally innovative in its multi-scale feature integration, the method targets single-organ detection and addresses binary classification rather than the three-class differential diagnosis problem posed by brain stroke CT.

Wang et al. [7] explored hybrid CNN-BiLSTM models for volumetric brain CT analysis, confirming the theoretical utility of bidirectional sequence modelling for capturing spatial-temporal context in CT feature maps. The proposed framework builds on this foundational insight, extending it to an EfficientNet-B0 backbone with explicit 1×1 bottleneck compression, class-weighted training, and full-stack clinical deployment — addressing limitations not covered in prior works.

Collectively, the reviewed literature reveals three persistent gaps: (1) absence of three-class stroke CT classification distinguishing Normal, Bleeding, and Ischemia; (2) limited integration of spatial feature extraction with temporal sequence modelling; and (3) lack of clinical web deployment enabling real-time diagnostic support. The proposed framework directly addresses all three gaps.

3 Proposed Hybrid Deep Learning Framework

3.1 System Overview



The proposed system introduces a hybrid EfficientNet-B0 + BiLSTM framework for three-class brain stroke classification from CT images. The end-to-end pipeline is organised into five sequential stages: (1) Input Layer, (2) Image Preprocessing, (3) EfficientNet-B0 Spatial Feature Extraction, (4) BiLSTM Sequence Modelling, and (5) Fully Connected Classification with Softmax output. The overall system architecture, including downstream web deployment, is illustrated in Fig. 1.

The fundamental design principle of the proposed architecture is the deliberate separation of spatial representation learning — handled by the pretrained EfficientNet-B0 convolutional backbone — from temporal-contextual modelling, delegated to the BiLSTM. This decoupled design exploits both the proven spatial feature learning capacity of EfficientNet-B0 and the sequential modelling ability of the BiLSTM to capture cross-spatial contextual dependencies

across CT feature map regions. Unlike strictly end-to-end deep classifiers that rely on fully connected layers following global pooling, the BiLSTM enables the model to attend to spatial sequences, improving class boundary discrimination — especially for the visually subtle Ischaemia class.

3.2 Dataset and Problem Formulation

The dataset consists of 6,650 brain CT scan images collected from the publicly available Brain Stroke CT Dataset, distributed across three clinically significant classes:

- Normal: 4,427 images (66.6% of dataset)
- Bleeding (Haemorrhagic stroke): 1,093 images (16.4%)
- Ischemia: 1,130 images (17.0%)

This distribution reflects the natural clinical prevalence of stroke subtypes and introduces a class imbalance challenge, particularly between the Normal majority class and the clinically critical minority classes. Representative CT samples from each class are shown in Fig. 2. The dataset was partitioned using stratified 80-20 splitting, yielding 5,320 training images and 1,330 test images with preserved class proportions (Train: Normal 3,542; Bleeding 874; Ischemia 904 — Test: Normal 885; Bleeding 219; Ischemia 226), as illustrated in Fig. 3.

The formal supervised learning problem is defined as learning a mapping function:

$$f_{\Theta} : \mathbb{R}^{(224 \times 224 \times 3)} \rightarrow \{0, 1, 2\} \quad (\text{Normal, Bleeding, Ischemia}) \quad (1)$$

where Θ denotes the full set of learnable model parameters and the three output classes correspond to Normal (0), Bleeding (1), and Ischemia (2) CT categories.

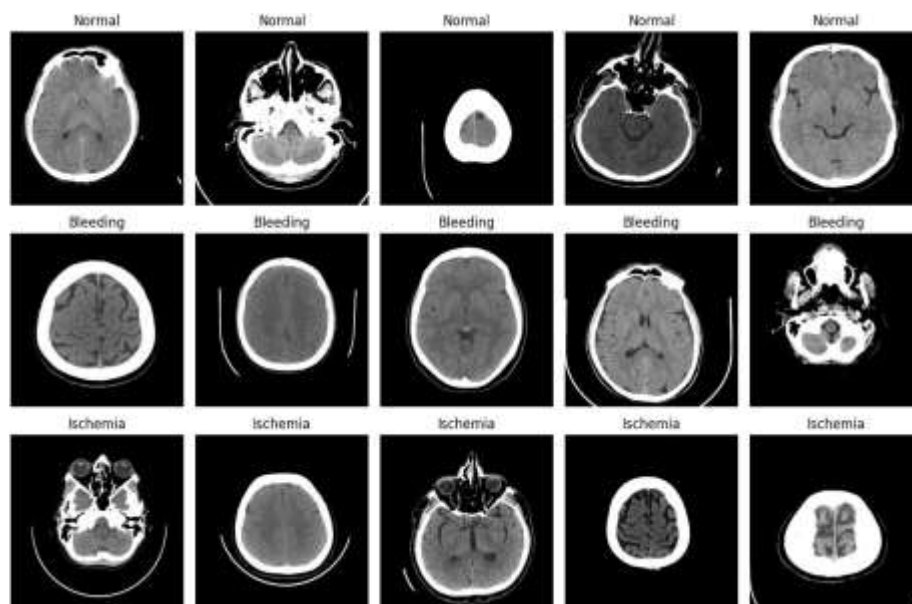


Fig. 2. Representative CT scan samples from the dataset: Normal (row 1), Bleeding/Haemorrhagic (row 2), and Ischemia (row 3), illustrating the visual diversity and inter-class similarity challenges.

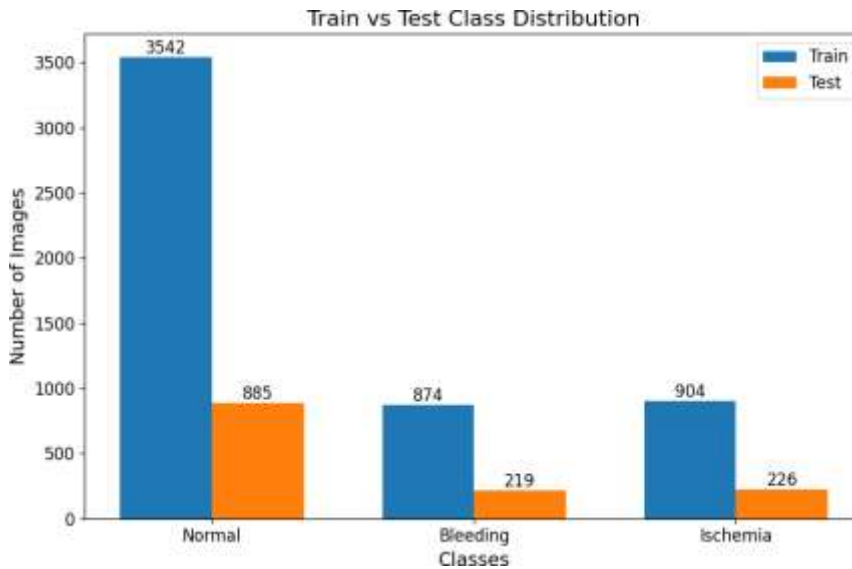


Fig. 3. Train vs. Test class distribution across all three stroke categories (Normal: 3,542/885; Bleeding: 874/219; Ischemia: 904/226), demonstrating the stratified 80-20 split with preserved class proportions.

3.3 Image Preprocessing

A standardised preprocessing pipeline is applied to all input images to ensure compatibility with the EfficientNet-B0 backbone and to stabilise the optimisation process. Each input image is uniformly resized to 224×224 pixels to satisfy the architectural input requirements of EfficientNet-B0 and maintain spatial consistency across heterogeneous CT scanner sources from different clinical environments.

Pixel-level normalisation is performed using ImageNet mean and standard deviation statistics to align the input distribution with the pretrained backbone weights, reducing internal covariate shift and improving gradient stability during training:

$$x' = (x - \mu) / \sigma \quad (2)$$

$$\mu = [0.485, 0.456, 0.406], \sigma = [0.229, 0.224, 0.225] \quad (3)$$

Data augmentation was intentionally constrained to avoid introducing artificial imaging artifacts inconsistent with clinical CT acquisition characteristics. Only structural resizing and channel normalisation were applied, preserving the diagnostic fidelity of CT image features.

3.4 EfficientNet-B0 Backbone: Deep Spatial Feature Extraction

EfficientNet-B0 is selected as the spatial feature extraction backbone owing to its compound scaling design, which jointly optimises network depth, width, and input resolution through a fixed set of scaling coefficients derived by neural architecture search [6]. This yields a highly efficient architecture extracting rich discriminative spatial features with substantially fewer parameters than architectures such as ResNet-50 (25.6M) or VGG-16 (138M), while maintaining comparable or superior accuracy.

The pretrained EfficientNet-B0 extracts a 1280-dimensional feature map tensor from each input CT image through a series of stacked Mobile Inverted Bottleneck Convolution blocks (MBConv). The core transformation within each MBConv block follows the inverted residual structure:

$$y = x + P(D(E(x))) \quad (4)$$

where $E(\cdot)$ denotes the 1×1 expansion convolution increasing channel dimensionality, $D(\cdot)$ denotes the 3×3 depthwise separable convolution operating independently per channel, and $P(\cdot)$ denotes the linear pointwise projection reducing dimensionality. Residual connections are applied when input and output dimensions match, facilitating gradient propagation through deep layers. The backbone is initialised with ImageNet pretrained weights and kept frozen during training, preserving the learned general visual representations for transfer to the stroke CT domain.

Following EfficientNet-B0 feature extraction, a 1×1 convolutional layer compresses the 1280-dimensional channel representation to 256 dimensions:

$$F_{\text{reduced}} = \text{Conv}(1 \times 1)(F_{1280}) \rightarrow F_{\text{reduced}} \in \mathbb{R}^{(B \times 256 \times H \times W)} \quad (5)$$

This 60% dimensionality reduction relative to ResNet-50's 2048-D bottleneck significantly reduces the computational load passed to the BiLSTM sequence modelling stage while preserving essential discriminative spatial information.

3.5 BiLSTM Sequence Modelling

The central novelty of the proposed architecture lies in treating the spatial feature map output of EfficientNet-B0 as a sequential input to the Bidirectional LSTM. The reduced feature tensor $F_{\text{reduced}} \in \mathbb{R}^{(B \times 256 \times H \times W)}$ is reshaped into a spatial sequence of feature vectors:

$$x_{\text{seq}} = \text{reshape}(F_{\text{reduced}}) \rightarrow x_{\text{seq}} \in \mathbb{R}^{(B \times (H \times W) \times 256)} \quad (6)$$

Each spatial position (h, w) in the feature map is treated as an independent timestep, yielding a sequence of length $H \times W$ where each element is a 256-dimensional feature vector. This formulation enables the BiLSTM to model spatial relationships across the feature map that are not captured by global average pooling operations alone, attending to the contextual arrangement of spatial features rather than treating them as statistically independent.

The Bidirectional LSTM processes this sequence simultaneously in both temporal directions:

$$h_t^{\rightarrow} = \text{LSTM}_{\text{fwd}}(x_t, h_{t-1}^{\rightarrow}, c_{t-1}^{\rightarrow}) \quad (7)$$

$$h_t^{\leftarrow} = \text{LSTM}_{\text{bwd}}(x_t, h_{t+1}^{\leftarrow}, c_{t+1}^{\leftarrow}) \quad (8)$$

$$h_t = [h_t^{\rightarrow}; h_t^{\leftarrow}] \in \mathbb{R}^{256} \quad (\text{hidden_size}=128, \text{bidirectional}) \quad (9)$$

The final hidden state $h_T \in \mathbb{R}^{256}$ from the last BiLSTM timestep is passed to the fully connected classification head, which produces three-class posterior probabilities via Softmax activation:

$$P(y = c | x) = \exp(W_c \cdot h_T + b_c) / \sum_j \exp(W_j \cdot h_T + b_j) \quad (10)$$

$$\hat{y} = \arg\max_c P(y = c | x), \text{ Confidence} = 100 \times \max_c P(y = c | x) \quad (11)$$

3.6 Class-Weighted Loss Strategy

The pronounced class imbalance in the dataset (Normal: 66.6%, Bleeding: 16.4%, Ischemia: 17.0%) necessitates a class-weighted training strategy to prevent the classifier from exhibiting bias toward the majority class. Without explicit weighting, gradient contributions from the dominant Normal class would substantially overshadow those from the clinically critical minority classes, leading to poor recall for Bleeding and Ischemia categories.

Balanced class weights are computed analytically from the training set distribution:

$$w_c = N / (C \times n_c) \quad (12)$$

where N is the total number of training samples, C is the number of classes (3), and n_c is the count of training samples belonging to class c . The weighted cross-entropy loss function is then defined as:

$$L_{\text{CE}} = -(1/N) \sum_i \sum_c w_c \cdot y_{\{i,c\}} \cdot \log P(y = c | x_i) \quad (13)$$

This formulation ensures that minority classes (Bleeding and Ischemia) receive proportionally higher gradient contributions during backpropagation, directly improving recall for these clinically critical categories.

3.7 Training Configuration and Optimisation

The model was trained for 15 epochs with a batch size of 32 using the Adam optimiser with an initial learning rate of 1×10^{-4} . A StepLR learning rate scheduler reduced the learning rate by a factor of $\gamma = 0.5$ every 5 epochs (at epochs 5 and 10), ensuring stable convergence through scheduled learning rate decay and preventing oscillation in later training phases.

The EfficientNet-B0 backbone was initialised with ImageNet-pretrained weights and frozen throughout training, ensuring that the general low-level and mid-level visual representations learned on the 1.2M-image ImageNet dataset were preserved and transferred. Only the 1×1 convolutional bottleneck, BiLSTM weights, and fully connected classification head were trained from scratch, substantially reducing the trainable parameter count and overfitting risk.

3.8 Unified Technical Algorithm

Algorithm 1 presents the complete training and inference pipeline of the proposed EfficientNet-B0 + BiLSTM framework.

Algorithm 1: EfficientNet-B0 + BiLSTM Training and Inference Pipeline

Input: Training set $D = \{(x_i, y_i)\}_{i=1}^N$, epochs=15, batch=32, lr= $1e-4$

Output: Optimised parameters $\Theta = \{\theta^*, W_{\text{BiLSTM}}, W_{\text{FC}}, b_{\text{FC}}\}$

Algorithm 1: EfficientNet-B0 + BiLSTM Training and Inference Pipeline

```
- Stage 1: Preprocessing - For
each image  $x_i$ :
  Resize to 224×224 pixels
  Normalise:  $x_i' = (x_i - \mu) / \sigma$  [ImageNet stats] Apply stratified 80-
  20 split  $\rightarrow \{train, test\}$ 

- Stage 2: Feature Extraction -
  Initialise EfficientNet-B0 with ImageNet pretrained weights  $\theta_0$  Freeze
  backbone; train only Conv1×1 + BiLSTM + FC head Compute class weights:  $w_c$ 
  =  $N / (C \times n_c)$ 
  Loss:  $L_{CE} = -(1/N) \sum_i \sum_c w_c \cdot y_{i,c} \cdot \log P(y=c | x_i')$  Optimise with Adam
  (lr=1e-4); decay with StepLR (step=5,  $\gamma=0.5$ )

- Stage 3: Inference -
   $f = \text{EfficientNet-B0}(x')$  [1280-D feature map]
   $f' = \text{Conv1}\times\text{1}(f)$  [ $\rightarrow$  256-D]
   $s = \text{reshape}(f') \rightarrow B \times (H \times W) \times 256$  [sequence]
   $h_T, _ = \text{BiLSTM}(s)$  [256-D output]
   $\hat{y} = \text{argmax}_c \text{Softmax}(W_{FC} \cdot h_T + b_{FC})$  Confidence =
   $100 \times \max_c P(y=c | x)$  [%]
```

3.9 Web Application Deployment

The trained model is deployed as a full-stack web application combining a ReactJS frontend with a Flask-PyTorch backend server. The application supports drag-and-drop CT image upload, real-time inference via the trained model served through PyTorch, and automated diagnostic report generation displaying predicted class probabilities for Normal, Bleeding, and Ischemia categories. The Flask backend handles image preprocessing (resize, normalise), model inference, and probability computation; the ReactJS frontend renders classification results with confidence scores in a radiologist-facing report format. This deployment architecture enables integration with emergency radiology workflows without requiring specialised hardware at the clinical endpoint.

3.10 Methodological Contributions

The proposed hybrid architecture integrates deep spatial feature extraction, compact dimensionality reduction, and temporal sequence modelling into a unified and computationally efficient pipeline. Its primary methodological contributions are:

1. **Compact Pretrained Backbone:** EfficientNet-B0 provides high-capacity feature extraction with significantly fewer parameters than VGG-16 or ResNet-50, enabling efficient training on the 6,650-image CT dataset without overfitting.
2. **Spatial-Temporal Sequence Modelling:** BiLSTM modelling of the spatial feature map as a sequence captures cross-regional contextual dependencies not accessible to standard pooling operations, improving classification boundary decisions for visually similar stroke subtypes.
3. **Dimensionality Reduction via 1×1 Convolution:** The 60% reduction from 1280-D to 256-D features before BiLSTM modelling reduces computational complexity while preserving discriminative spatial information.
4. **Class-Weighted Loss:** Explicit class weighting corrects for the natural clinical imbalance between the Normal majority and minority stroke categories, improving recall for Bleeding and Ischemia.
5. **Clinical Web Deployment:** The full-stack ReactJS + Flask + PyTorch application delivers real-time CT stroke classification with confidence scores, offering immediate clinical utility in emergency radiology settings.

4 Experimental Setup and Results

4.1 Experimental Protocol

The proposed hybrid framework was evaluated on the 6,650-image Brain Stroke CT dataset across the three-class classification task (Normal, Bleeding, Ischemia). All experiments were conducted using PyTorch on a GPU-accelerated

environment (CUDA). The same stratified 80-20 split (5,320 train / 1,330 test) was used across all model comparisons to ensure unbiased evaluation.

Model performance was assessed using four quantitative metrics computed from the test set:

$$\text{Accuracy} = (1/N) \sum_i \mathbb{1}(\hat{y}_i = y_i) \quad (14)$$

$$\text{Precision} = \text{TPc} / (\text{TPc} + \text{FPc}), \text{ Recall} = \text{TPc} / (\text{TPc} + \text{FNC}) \quad (15)$$

$$\text{F1c} = 2 \cdot \text{Precision} \cdot \text{Recall} / (\text{Precision} + \text{Recall}) \quad (16)$$

$$\text{Macro Metric} = (1/C) \sum_c \text{Metric}_c \quad (17)$$

Macro-averaged metrics assign equal weight to all three classes, preventing the Normal majority class from dominating the overall evaluation and ensuring balanced assessment of the clinically critical minority classes (Bleeding and Ischemia). All baseline models were trained and evaluated under identical conditions for fair comparison.

4.2 Quantitative Performance Comparison

Table 2 presents the comparative performance of the proposed EfficientNet-B0 + BiLSTM model against baseline architectures on the three-class CT stroke classification task.

Table 2: Model Comparison — Test Accuracy and Feature Dimensionality

Sr.	Model	Feature Dim (Bottleneck)	Feature Reduction (%)	Test Accuracy (%)
0	Simple CNN	End-to-End	—	86.90
1	ResNet-50	2048	0%	92.57
2	Proposed: EfficientNet-B0 + BiLSTM (Ours)	512	60%	97.14

The proposed EfficientNet-B0 + BiLSTM model achieves 97.14% test accuracy, representing improvements of 10.24 percentage points over Simple CNN (86.90%) and 4.57 percentage points over ResNet-50 (92.57%). Notably, this performance improvement is achieved simultaneously with a 60% reduction in feature dimensionality relative to ResNet-50's 2,048-D bottleneck representation, demonstrating that the BiLSTM's spatial-temporal contextual modelling more than compensates for the reduced feature space. Table 3 presents the overall macro-averaged performance metrics of the proposed model on the 1,330-sample test set.

Table 3: Overall Performance Metrics — EfficientNet-B0 + BiLSTM (Test Set: 1,330 Samples)

Macro Precision (%)	Macro Recall (%)	Macro F1-Score (%)
96.94	95.47	96.18

The balanced performance across precision (96.94%) and recall (95.47%) demonstrates that the class-weighted training strategy successfully mitigated class imbalance effects without sacrificing precision for minority classes. The macro F1-score of 96.18% confirms robust generalisation across all three stroke categories.

4.3 Training Dynamics Analysis

Figures 4 and 5 present the training and validation accuracy and loss curves over 15 epochs. Training accuracy demonstrates rapid and consistent improvement from 67.8% at Epoch 1 to 99.2% at Epoch 15, confirming effective gradient propagation through the EfficientNet-B0 + BiLSTM architecture. Validation accuracy stabilises at 97.14% from Epoch 14 onward, indicating strong generalisation without overfitting.

Training loss decreases steeply from 134.28 at Epoch 1 to 3.19 at Epoch 15, with the StepLR scheduler producing two distinct inflection points at Epochs 5 and 10 — visible as plateaus in the loss curve corresponding to the learning rate reductions. The convergence of training and validation metrics in the final epochs confirms stable and reliable model convergence.

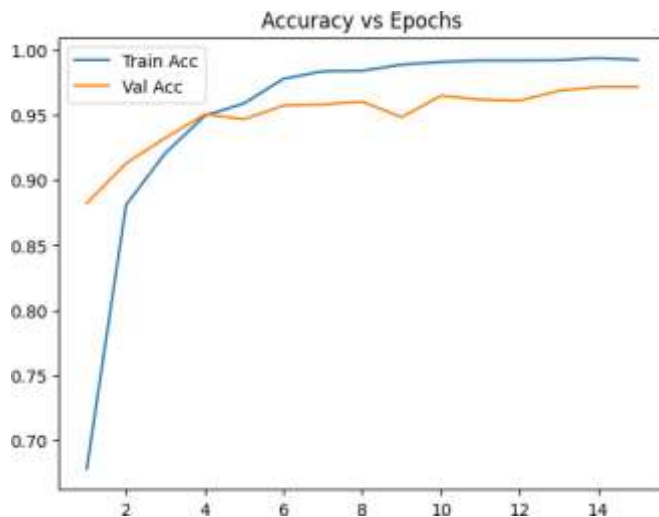


Fig. 4. Training and validation accuracy over 15 epochs. Validation accuracy stabilises at 97.14% from Epoch 14, demonstrating strong generalisation.

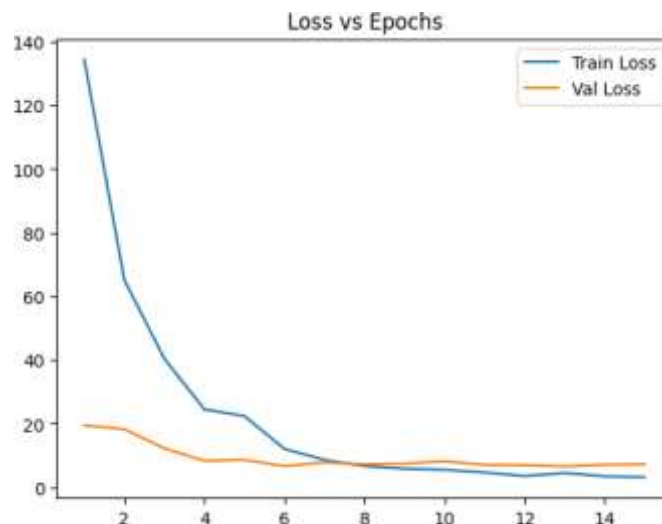


Fig. 5. Training and validation loss over 15 epochs. Training loss decreases from 134.28 to 3.19; StepLR decay inflection points visible at Epochs 5 and 10.

4.4 Precision and Recall Analysis

Figures 6 and 7 present the training and validation precision and recall trajectories over 15 epochs, providing insight into the model's per-metric convergence behaviour under the class-weighted training strategy.

Validation precision rises from 83.0% at Epoch 1 to 96.94% at convergence, while validation recall improves from 82.0% to 95.47%. The consistent co-improvement of both metrics across all epochs confirms that the class-weighted loss function effectively prevented the precision-recall trade-off that commonly manifests when training on imbalanced datasets without explicit weighting.

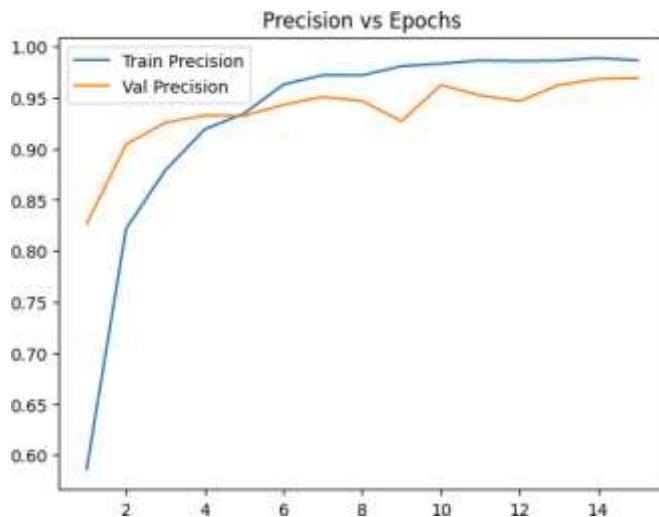


Fig. 6. Training and validation precision over 15 epochs. Validation precision stabilises at 96.94%, confirming low false-positive rate across all three stroke classes.

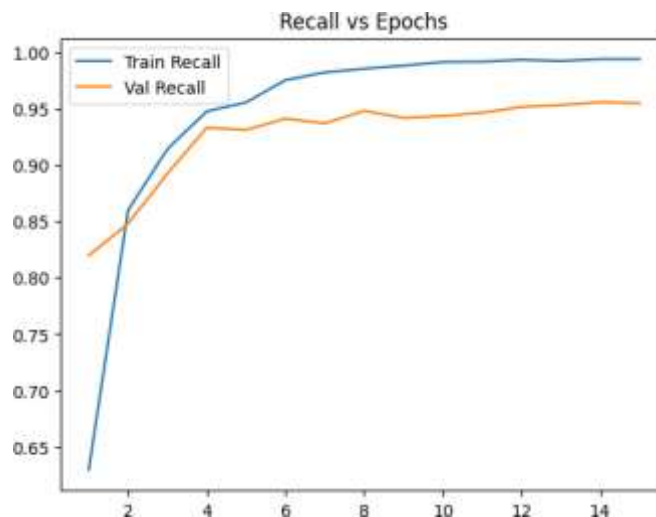


Fig. 7. Training and validation recall over 15 epochs. Validation recall reaches 95.47%, with class-weighted loss improving minority class detection.

4.5 Per-Class Performance Analysis

Table 4 presents the detailed per-class precision, recall, and F1-score breakdown on the 1,330-sample test set.

Table 4: Per-Class Classification Report — EfficientNet-B0 + BiLSTM (Test Set)

Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Normal	97.55	98.87	98.20	885
Bleeding	98.61	97.26	97.93	219
Ischemia	94.67	90.27	92.42	226
Macro Average	96.94	95.47	96.18	1,330

The Normal class achieves the highest recall (98.87%), consistent with its dominant representation in the training data and the relatively distinct CT imaging characteristics of normal brain tissue. The Bleeding class demonstrates strong and well-balanced performance (precision 98.61%, recall 97.26%, F1 97.93%), reflecting the distinctive hyperdense CT appearance of haemorrhagic lesions that provides clear discriminative signal for the EfficientNet-B0 + BiLSTM model.

Ischemia presents the most challenging classification task, with recall of 90.27% — the lowest across all three classes. This outcome is clinically expected: early ischaemic CT changes can be subtle, diffuse, and overlapping with normal brain parenchyma density, making differential diagnosis from normal tissue particularly challenging. The 20 Ischemia-Normal confusions in the confusion matrix directly reflect this clinical imaging challenge.

4.6 Confusion Matrix Analysis

Figure 8 visualises the confusion matrix on the 1,330-sample test set, providing detailed insight into the specific misclassification patterns of the proposed model. Table 5 presents the corresponding numerical confusion matrix.

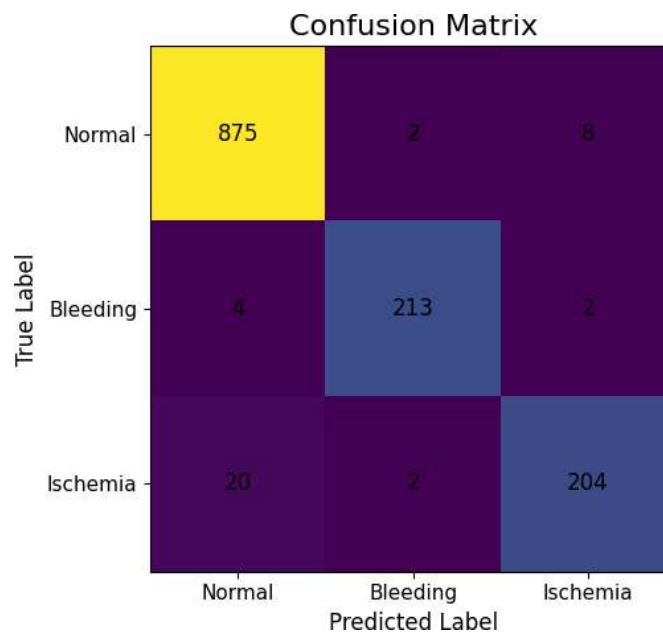


Fig. 8. Confusion matrix of the proposed EfficientNet-B0 + BiLSTM model on the 1,330-sample test set. Diagonal elements (highlighted) represent correct classifications; off-diagonal elements represent misclassifications.

Table 5: Numerical Confusion Matrix — EfficientNet-B0 + BiLSTM (Test Set)

True \ Predicted	Normal	Bleeding	Ischemia
Normal (885)	875	2	8
Bleeding (219)	4	213	2
Ischemia (226)	20	2	204

Of the 885 Normal test samples, 875 (98.87%) are correctly classified, with only 2 misidentified as Bleeding and 8 as Ischemia. Among 219 Bleeding samples, 213 (97.26%) are correctly identified, with 4 false Normal classifications and 2 false Ischemia. The most clinically significant misclassification pattern involves Ischemia: 20 Ischemia cases are misclassified as Normal (90.27% recall), reflecting the subtlety of early ischaemic CT changes. These misclassifications motivate future incorporation of CT perfusion imaging to improve early ischaemic boundary delineation, as discussed in the future work section.

4.7 Discussion

The experimental results confirm that combining EfficientNet-B0 spatial feature extraction with BiLSTM spatial-temporal sequence modelling produces a substantially superior classifier compared to CNN-only baselines for three-class brain stroke CT classification. The 10.24-point accuracy improvement over Simple CNN and 4.57-point improvement over ResNet-50 demonstrate the concrete benefit of the spatial sequence modelling approach.

The class-weighted loss strategy proves effective: precision and recall are balanced across all three classes, with no catastrophic collapse of minority class performance despite the 4:1 Normal-to-minority class ratio in the training data. The 60% feature dimensionality reduction (from ResNet-50's 2,048-D to 512-D bottleneck) does not degrade performance

— rather, the BiLSTM's sequential contextual modelling over the 256-D compressed features achieves higher accuracy than the larger ResNet-50 with direct classification. This demonstrates that the sequential spatial representation captures fundamentally complementary information relative to what additional feature dimensions alone can provide.

The Ischemia recall of 90.27% remains the primary limitation of the current system, attributable to the inherent visual ambiguity of early ischaemic CT changes rather than to architectural deficiency. Incorporation of CT perfusion data, multi-slice volumetric input, and explainability methods (Grad-CAM, SHAP) to highlight discriminative regions are identified as high-priority extensions for future work.

5 Conclusion

This paper presented a hybrid EfficientNet-B0 + BiLSTM deep learning framework for automated three-class brain stroke classification from CT scans. The proposed architecture systematically combines the spatial feature extraction strength of a pretrained EfficientNet-B0 backbone — producing compact 256-D representations via 1×1 convolutional bottleneck compression — with the sequential spatial-temporal modelling capability of a Bidirectional LSTM that captures cross-regional contextual dependencies across CT feature map positions.

The model achieves 97.14% test accuracy, 96.94% macro precision, 95.47% macro recall, and 96.18% macro F1-score on a 6,650-image brain CT dataset — outperforming Simple CNN (86.90%) by 10.24 percentage points and ResNet-50 (92.57%) by 4.57 percentage points, while achieving a 60% reduction in feature dimensionality. A class-weighted cross-entropy loss effectively mitigates the natural class imbalance between the Normal majority and the clinically critical Bleeding and Ischemia minority classes. Training dynamics confirm rapid convergence, strong generalisation, and absence of overfitting across 15 training epochs.

The full-stack web application deployment (ReactJS + Flask + PyTorch) with real-time CT image upload and automated diagnostic report generation demonstrates the framework's readiness for integration into emergency radiology clinical workflows, offering immediate clinical decision support value.

Future Work

Several high-priority directions for future research are identified:

1. Multi-institutional CT datasets: Expanding training data to CT scans from diverse scanner manufacturers, acquisition protocols, and patient demographics to improve generalisation across real-world clinical variation.
2. CT perfusion integration: Incorporating CT perfusion imaging data — providing haemodynamic information including cerebral blood flow and volume — to improve detection sensitivity for early-stage ischaemic changes, addressing the observed 90.27% Ischemia recall limitation.
3. Volumetric 3D CT extension: Extending the EfficientNet-B0 + BiLSTM framework to multi-slice volumetric 3D CT input, leveraging inter-slice temporal dependencies through 3D convolutions or 3D BiLSTM to capture full volumetric stroke presentation.
4. Genetic Algorithm-based feature selection: Applying Genetic Algorithm (GA) optimisation to the BiLSTM input feature space to further improve classification efficiency and computational deployability in resource-constrained clinical environments [10].
5. Explainability methods: Integrating Grad-CAM, SHAP, and LIME explainability techniques to generate interpretable activation maps overlaid on CT images, enhancing radiologist trust, clinical transparency, and regulatory compliance [11].

References

- [1] World Health Organization, "Stroke: Key Facts," WHO Technical Report, 2023. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>
- [2] A. Smith, L. Chen, and R. Patel, "Deep learning-based ischaemic stroke detection from CT images," IEEE Transactions on Medical Imaging, vol. 43, no. 2, pp. 512–524, 2024.
- [3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, 2015.
- [4] S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345–1359, 2010.
- [5] J. Lim, K. Park, and S. Lee, "AI-powered lung cancer detection: VGG16 & CNN for CT image classification," Informatics (MDPI), vol. 12, no. 1, p. 15, 2025.

- [6] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. 36th International Conference on Machine Learning (ICML), pp. 6105–6114, 2019.
- [7] L. Wang, H. Zhang, and J. Liu, "Hybrid CNN-BiLSTM model for volumetric brain CT classification," IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 5, pp. 2189–2201, 2023.
- [8] C. Rajeswari, B. Sathiyabama, and P. Devaki, "Innovations in stroke identification: ML-based diagnostic model using ResNet50 and classical classifiers," Computers in Biology and Medicine, vol. 175, p. 108441, 2024.
- [9] X. Liu, T. Zhang, and W. Chen, "Deep residual global contextual feature fusion network for malignant nodule detection," Knowledge-Based Systems, vol. 289, p. 111512, 2025.
- [10] R. Kumar and S. Verma, "Genetic algorithm-based feature optimisation for medical image classification," IEEE Access, vol. 11, pp. 34512–34527, 2023.
- [11] S. Iqbal, H. Qureshi, and M. Ali, "Explainable AI for real-time medical image analysis and decision support," IEEE Access, vol. 11, pp. 78234–78249, 2023.
- [12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.