Journal of Scientific & Industrial Research
Vol. 85, July 2026, pp. 610-620
DOI: 10.56042/jsir.v85i7.35619

PCA–GA Optimized CNN–LSTM for Multimodal Stroke Detection

Dipali Nilesh Dhake[1]*, Puja Gholap[2], Shital Y. Solanke[3], Vilas Shamrao Bugade[4], Sayali Kokane[5] & Ganesh H. Patil[6,7]*

[1]Department of Electronics and Telecommunication Engineering, PCET's Pimpri Chinchwad College of Engineering and Research, Ravet, Pune 412 101, Maharashtra, India

[2]Department of Computer Engineering, Sharadchandra Pawar College of Engineering, Junnar 412 409, Pune, Maharashtra, India

[3]Department of Mathematics, School of Science and Humanities, Ramdeobaba University, Ramdeobaba Tekdi, Katol Road, Nagpur 440 013, Maharashtra, India

[4]Department of Electrical Engineering, Marathwada Mitra Mandal's College of Engineering, Pune 411 052, Maharashtra, India

[5]Information Technology Department, ABMSP's Anantrao Pawar College of Engineering and Research, Parvati, Pune 411 009, Maharashtra, India

[6]Department of Electrical Engineering, Dr. D. Y. Patil Institute of Technology, Sant Tukaram Nagar, Pimpri, Pune, 411 018, Maharashtra, India

[7]Dnyaan Prasad Global University (DPGU), School of Technology and Research, Dr. D. Y. Patil Unitech Society, Sant Tukaram Nagar, Pimpri, Pune 411 018, Maharashtra, India

Received 20 August 2026; revised 26 September 2026; accepted 30 September 2026

Stroke is a major cause of mortality and disability, making timely and reliable detection important for clinical management. Multimodal analysis integrating Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and clinical information can provide complementary diagnostic information, but the fusion of heterogeneous data may introduce correlated and redundant features, increasing computational complexity. This study proposes a PCA–GA optimized CNN–LSTM framework for automated stroke detection using multimodal CT, MRI, and clinical data. CT and MRI data were subjected to normalization, noise reduction, artifact correction, and feature extraction, while clinical variables were standardized. The extracted features were fused and processed using Principal Component Analysis (PCA) to reduce dimensionality, followed by Genetic Algorithm (GA)-based feature selection to identify informative features. The optimized feature representation was classified using a CNN–LSTM architecture comprising four convolutional blocks and two stacked LSTM layers. A patient-level 80:20 training–testing split with stratified 10-fold cross-validation was employed to reduce information leakage. On the independent test set, the proposed framework achieved 98.2% accuracy, 98.01% precision, 98.4% recall, 98.2% F1-score, 98.0% specificity, and an Area Under the ROC Curve (AUC) of 0.98. Compared with SVM, Random Forest, and CNN baseline models evaluated using the same dataset and protocol, the proposed framework showed higher performance across the reported metrics. The findings demonstrate the potential of combining PCA, GA, CNN, and LSTM for efficient multimodal stroke classification and clinical decision support. Further external multi-site validation and explainable AI methods are required to assess generalizability and interpretability.

Keywords: Artificial intelligence, Deep learning, Feature medical imaging, Genetic algorithm, Healthcare analytics

Introduction

Stroke is a prominent neurodegenerative disease and one of the top killers and causes of disabilities across nations. It is due to ischemic occlusion or haemorrhagic damage to cerebral blood vessels, and a delayed diagnosis may culminate in irreversible brain injury, disability and death. Early diagnosis of stroke has clinical implications as it facilitates timely treatment, which can affect the outcome of patients. Conventional diagnostic approaches involve clinical evaluation and imaging modalities (such As Computed Tomography (CT) and Magnetic Resonance Imaging (MRI)), which depend on the experience of specialists, especially vulnerable to discrepancy in decision-making.[1–3] The increasing availability of imaging and clinical information has inspired much research in the field of automated stroke detection from CT scans using Artificial Intelligence (AI) and deep learning methods.[4,5]

Incorporating multimodal methods with a combination of clinical parameters, CT, and MRI may give some additional insight as well as an improved characterization of the features associated with stroke.[6] On the other hand, however heterogeneous modalities

*Authors for Correspondence
E-mail: ganeshkochur@gmail.com, dipalippatil@gmail.com

are not all features and thus jointly induces high dimensional input data with correlated and duplicated information hence may lead to substantial increase in processing workload. These models typically operate on separate feature extraction or classification modules over each modality and fuse the outputs, which gives rise to some feature redundancy and computational burden. These limitations relate especially to clinical decision-support systems that need to be processed with optimal time and predictable times. This necessitates a comprehensive methodology wherein redundancy is reduced to only few features which consume less computational constraint while preserving multimodal information.

The objectives of this study are to:

1. Develop a hybrid CNN–LSTM framework for automated stroke detection using multimodal CT, MRI, and clinical data.
2. Reduce feature dimensionality and redundancy in the fused multimodal data using Principal Component Analysis (PCA).
3. Select the most informative features from the reduced feature space using a Genetic Algorithm (GA).
4. Evaluate the performance of the PCA–GA optimized CNN–LSTM framework using accuracy, precision, recall, F1-score, specificity, ROC-AUC, and confusion matrix analysis.
5. Compare the proposed framework with conventional machine learning and deep learning models under the same experimental and evaluation conditions.

Multimodal prediction of stroke detection in due to CT, MRI and clinical parameters can elicit complementary diagnostic information; however, the integration of heterogenous data posits high-dimensional feature spaces that harbor both correlated and redundant information. This is an added burden computationally and may reduce the usefulness of clinical decision-support systems. If separate feature extractors and classifier are performed for different modality this obviously creates separation of redundancy existing in the approaches which are already available. Therefore, a joint framework is required to simultaneously gain the most informative multimodal features while ignoring redundancy and duplicate information.

Review of Literature

The advancements in artificial intelligence provide revolutionary benefits for neurological disorder diagnosis, particularly stroke. Numerous machine learning and deep learning approaches have been applied to the analysis of clinical and imaging datasets. Convolutional neural networks are suitable for working with CT and MRI scans in transfer learning models (VGG16, ResNet or AlexNet) show good classification results. Ensemble learning techniques improve the accuracy of predictions. Feature selection methods are techniques that decrease the number of variables under consideration (e.g. Principal Component Analysis, Genetic Algorithms (GA), Particle Swarm Optimization (PSO) impacting directly on computational cost and making models more efficient). But the problems with data arbitrary variability, high computational costs and interpretability reveal that Hybrid mode is still in need of more optimization by feature selection. Deep neural network and transfer learning methods have been previously successful for stroke classification.[7,8]

Integration of different imaging platforms with multimodal imaging modalities increases the diagnostic accuracy compared to a mono-modal approach.[9] Hybrid models combine the fine local feature extraction capability of VGG 16 with context learning ability of Vision Transformers to detect stroke pattern accurately.[10,11] Transformer models in the global attention mechanism helps you figure out where to focus your mind when looking at Image.[12–14]

Another study shows that a combination of architecture such as VGG-16 with other machine learning methods can exceed 98% accuracy, which is higher than any individual model.[15] Although transformer attention mechanisms locate the most important regions related to stroke and self-attention provide satisfactory explanation for high-order visual relationships, this study could offer numerous observations on simple, intuitive morphological artery features.[16] This way, both local and the global features can be obtained similar to clinical diagnostic reasoning.[17] Transformers and networks like U-Net are more informative for feature representation since learning carries semantic information through the down-sampling.[18] Using multi-head attention mechanisms, hybrid models perform better in segmentation tasks by concentrating on the correct parts of the input while disregarding data points that are placed incorrectly.[19] Enabling comprehensive feature extraction and detecting little abnormalities and major pathologies because few size ranges may appear as lesions.[20,21]

Scientifically, this study is significant because it integrates the multimodal stroke data (CT, MRI, and clinical features) with a two-stage feature optimization strategy using PCA and GA, followed by classification via hybrid CNN–LSTM architecture. Existing approaches usually focus on either single-modal data or operate on a single-deep model or feature selection approach independently; The proposed framework is an end-to-end multi-level deep feature extraction framework using spatiotemporal characteristics and produce reduced features but improve computational efficiency in stroke-related data processing. This design can provide more stable and clinically applicable analysis results based on the integration of these two clinical variables, which is innovative.

Multimodal stroke studies tend to mix imaging and clinical data and hybrid deep-learning models have enhanced feature representation across different modalities. However, all these reviewed methods mostly consider either multimodal fusion or deep learning classification while no approach combines multimodal feature fusion, PCA+GA-based feature selection, and CNN–LSTM classification in a single pipeline. Therefore, this research fills the gap between an integrated framework that limits redundancy of fused CT, MRI and clinical features before deep-feature learning in both spatial and sequential domains. The proposed study fills the gap by multi-step multimodal feature fusion → PCA → GA → CNN → LSTM→Stroke classification.

The proposed framework is novel compared to existing multimodal stroke detection methods in the way that it sequentially combines multimodal feature-level fusion, PCA-based dimensionality reduction, GA-based feature selection, and CNN–LSTM-based spatial–sequential learning into one classification pipeline. Principal Component Analysis (PCA) is employed to discard the correlated and redundant features of the fused CT, MRI and clinical feature space, while Genetic Algorithm (GA) anchors all informative reduced features which need pre-process before deep learning classification. The CNN layers learn spatial characteristics of the optimized feature representation while LSTM layers model interdependencies among the features obtained from these CNN layers. Consequently, the proposed one is fundamentally different from approaches that predominantly pursue direct multimodal fusion, single feature-selection strategies or separate deep learning structures by integrating redundancy reduction, feature optimization, spatial learning and sequential features learning into a unified framework.

Theoretical Framework of Deep Learning Architectures

Convolutional Neural Networks (CNNs)

This property (convolutions can mine hierarchical spatial features from an image) makes CNNs so popular for medical image analysis. As noted earlier, the first few convolutional layers often learn to recognize low level features, whereas deeper higher-level appropriate feature representations capture more complex anatomical and pathological patterns. For histology and MRI images, it utilized CNN-based models for feature extraction. It is proposed to construct CNN layers as the first principal spatial feature mapping component that would then be processed by LSTM networks, thus allowing for this sequential input of information.

VGG16 and ResNet

The VGG16 model is an architecture consisting of just a stack of small 3×3 filters and pooling operations. As its architecture is very simple, it has been widely used for transfer-learning model for medical imaging classification. ResNet introduced the main novelty of using residual connections that let information skip across many layers in a lateral manner while minimizing some of the optimization issues that were related to deeper networks. Many of these architectures have been studied in the medical literature for both classification tasks and stroke related image analysis. They are especially relevant to the current study in that they can serve as canonical deep CNN architectures against which hybrid approaches might be compared conceptually.

U-Net

U-Net is a convolutional neural network architecture for efficient biomedical image segmentation, which follows an encoder–decoder topology consisting of skip connections. It consists of an encoder that captures high-level features and a decoder that reconstructs spatial information in a step-by-step manner to generate segmentation maps that are visually meaningful. For U-net & variants, these are often beneficial when an image must be label then for stroke lesions (though its often called segmented too), is the image either of a person with some strikes occurring or not. Even though the present study mainly regards

Table 1 — Theoretical framework of deep learning architectures

Model	Core approach	Stroke-related application
CNN	Extracts spatial patterns through convolution layers	CT/MRI image feature extraction
VGG16	Uses multiple sequential convolution layers	Transfer-learning-based stroke classification
ResNet	Employs residual connections to improve deep learning	Medical image classification
U-Net	Combines encoder–decoder layers with skip connections	Stroke lesion segmentation
Vision Transformer	Uses self-attention to model image relationships	Analysis of multiple brain-image regions
Attention	Assigns greater weight to important features	Relevant feature identification and interpretation
CNN–LSTM	Integrates CNN spatial features with LSTM sequence learning	Multimodal stroke classification

classification, U-Net-based approaches can be considered as a conceptual basis for what spatial patterns must be preserved in medical imaging data.

Vision Transformers and Attention Mechanisms

Self-attention in transformer-based architecture allows for relationships between different areas of input to be modelled rather than relying on solely local convolutional operations. Vision Transformers (ViTs) segment images into patches and leverage attention mechanisms to learn relationships among the patches. Attention mechanisms provide the ability to focus on diagnostically relevant observations and serve as a mechanism to model long-range dependencies. The recent stroke-related studies have investigated the applicability of Transformer-based and hybrid architectures for classification, segmentation, and prognosis.

Vision Transformers and Attention Mechanisms

CNN–LSTM An ensemble of CNN and LSTM layers where the CNN layer better captures spatial representation and the LSTM capture temporal correlation. The CNN layers extract informative spatial features and the LSTM layers model dependencies within the created feature sequence. So, in the proposed framework, this combination follows PCA-based dimensionality reduction and GA based feature selection. Thus, the proposed model is based on a theory that integrates multiple aspects including multimodal feature representation, dimension reduction, optimal features selection, spatial characteristic extraction and sequential characteristics learning. The architecture implemented here is consists of four convolutional blocks followed by two LSTM layers with 128 and 64 hidden units, respectively. The comparison between Theoretical Framework of Deep Learning Architectures is explained in Table 1.

Proposed Methodology

Multimodal Feature Fusion and PCA–Ga Feature Optimization

A feature-level multimodal fusion mechanism to fuse imaging and clinical (tabular) information is proposed in the devised framework. The CT and MRI data were initially pre-processed, including image normalization, noise reduction, and artifact correction; clinical variables such as hypertension, diabetes mellitus, smoking history, cardiovascular disease, blood pressure and cholesterol were standardized. The corresponding features extracted from the imaging data are then combined with the standardized clinical features to create a combined multimodal feature representation for each subject. This level of feature fusion enables complementary information from both imaging and clinical data to be analysed simultaneously in a unified classification framework.

The extracted multimodal feature space is then subjected to Principal Component Analysis (PCA) which discards redundant and highly correlated information to reduce the dimensionality of the data. PCA can be thought of as a dimensionality reduction technique that reduces the dimensionality of the combined feature representation into a smaller number of principal components, while still preserving most of the essential information from the original multimodal data. Thus, after PCA, Genetic Algorithm (GA) is applied on the reduced feature space to select those principal components which are most discriminative for classification. So PCA is actually applying dimensionality reduction, whereas GA applies feature selection. The CNN–LSTM model takes in the optimized feature representation for classification tasks. Thus, the whole processing pipeline is CT/MRI feature extraction and clinical feature preparation → multimodal feature fusion → PCA dimension reduction → GA feature selection →

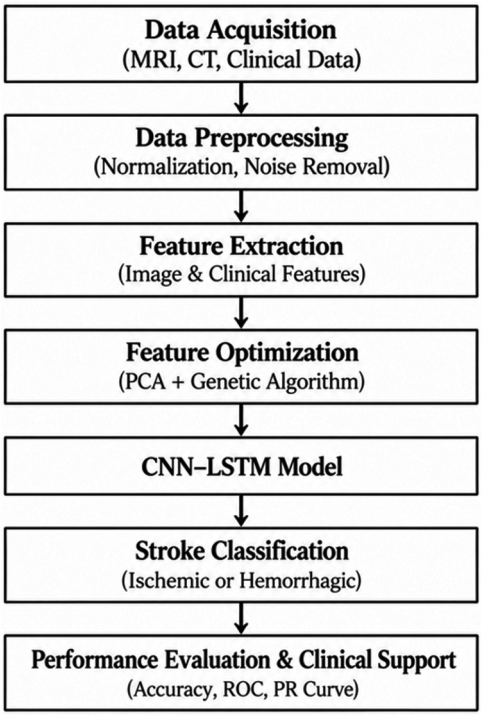


Fig. 1 — System architecture

optimized representation of features → CNN–LSTM classification. It also specifies the sequence used to delineate where PCA and GA have their definition of the feature space, just before CNN–LSTM classification.

The proposed workflow which consists of collection of all available data (MRI, CT and clinical), followed by preprocessing, feature extraction and finally optimization through PCA and Genetic Algorithms has explained in Fig. 1. These optimized features are utilized to classify stroke and evaluate performance via a hybrid CNN–LSTM model to support clinical decision-making. This section also specifies the mathematical formulations for pre-processing, feature selection, dimensionality reduction, model training, and evaluation. Normalization ensures uniform data scaling, and PCA enhances feature representation. CNN and LSTM Models Learn Spatial and Temporal Patterns in Biomedical Data. Accurate classification and reliable validation are both supported by loss functions and evaluation metrics.

Dataset Description and Ethical Considerations

These datasets were chosen to give valuable data on stroke imaging and clinical details. The imaging data were either from CT and MRI (CT, MRI-DWI, and FLAIR sequences). Records corresponded to the same patient if they were from the same dataset, since records referred to independent anonymized cohorts across both source datasets. Patients were not identified or matched cross-dataset. CT, MRI and clinical variables were combined only when the information was available from the same subject in the original source dataset or cohort. This process avoided the imitative combination of imaging and clinical data from unrelated patients.

CT, MRI and clinical data underwent modality-specific preprocessing prior to extraction and multimodal fusion. CT images standardization to the same spatial resolution followed by intensity normalization, noise reduction, and artifact correction were done on CT images to improve dataset consistency. MRI data (DWI, and FLAIR sequences) underwent the same standardization process for both resolution normalization, intensity normalization, noise reduction, and imaging artifact correction. Such adjustments were made to reduce variations coming from different sources of in vivo images and production conditions. In the case of clinical data, missing/invalid records were either dropped or modified for each of the selected variables (hypertension, diabetes mellitus, smoking history, Cardiovascular Disease (CVD), blood pressure and cholesterol all statistically significant on univariate analysis); these were then normalized with respect to the imaging-derived features. Prior to developing a model, duplicate, incomplete, poor quality and diagnostically ambiguous records were excluded. These processed CT features, MRI features, and clinical information were then merged at the feature level to form the unified multimodal representation for downstream PCA based dimensionality reduction followed by GA-based feature selection.

Records were excluded if they were duplicate, incomplete, of poor-quality or diagnostically ambiguous prior to model development. Study Population Participants with clinically and radiologically confirmed stroke (cerebral haemorrhage, or partially- or fully recovered ischemia) were compared with control subjects without cerebrovascular disease. Patients: The mean age of the selected subjects was 61.4 ± 12.3 years; males accounted for 52% and females for 48%. Other features were clinical such as hypertension, diabetes mellitus, smoking status, cardiac disease and blood pressure and cholesterol values. Imaging data was pre-processed with standardization of image resolution, intensity

Table 2 — Summary of datasets used for multimodal stroke analysis and their public availability

Dataset	Data Type / Modality	Sample Size	Public Access
ISLES 2016	MRI + clinical data	Not specified	Publicly available
ISLES 2017	Multispectral MRI + clinical data	51 patients	Publicly available
ISLES 2018	CT perfusion + MRI/DWI	103 patients	Publicly available
ISLES 2022	Multicentre MRI	400 subjects	Zenodo, CC BY 4.0
ISLES 2024	CT + MRI + clinical outcomes	245 patients	Zenodo
ATLAS v2.0	T1-weighted MRI + lesion masks	955 subjects; 1,271 including hidden data	Public release
CQ500	Head CT + radiologist labels	≈491 examinations	Publicly available
RSNA ICH	Head CT + hemorrhage labels	>25,000 examinations	Publicly available
AISD	NCCT + DWI reference	397 scans	Publicly available
APIS	Paired NCCT + ADC/MRI	Not specified	Public; registration required

normalization, noise reduction and artifact correction aimed at improving consistency across the source datasets.

The entire dataset was split at patient-level into 80% training and 20% independent testing data. No patient or imaging record was allowed to enter both subsets. The 80% training data were subsequently assessed through stratified 10-fold cross-validation, where the training data were split into ten folds according to stroke and non-stroke case class to maintain its distribution. In all iterations, nine folds were used for model development, and one-fold was set apart for validation. Significantly, all surrounding optimization processes were realized solely on the data within the respective training folds, both for PCA-based dimensionality reduction and GA-guided feature selection. The PCA transformation parameters were computed from the training segment and then applied to the paired validation fold, while GA feature selection was optimized with only data from the training set. Features were optimized without using the validation fold. The optimal pipeline, chosen through the cross-validation procedure, was tested on a 20% independent test set that was not touched at any step of training, PCA calculation, GA optimization and model selection. This patient-level protocol was used to reduce information leakage and an unbiased evaluation for the generalization capability of the PCA–GA optimized CNN–LSTM framework proposed.

The ten datasets used in the study were ISLES 2016, ISLES 2017, ISLES 2018, ISLES 2022, ISLES 2024, ATLAS, CQ500, RSNA, AISD, and APIS as shown in Table 2. The datasets were used as independent publicly available resources, and the analysis was performed on their available anonymized or de-identified records. No identifiable patient information or private hospital database records were accessed or used in this study. The dataset was used in an aggregated and non-identifiable form, and no personally identifying information was included in the analysis. As the present research involved secondary analysis of publicly available anonymized datasets and did not involve direct patient recruitment or interaction, no new informed consent was obtained from participants. The datasets were used in accordance with the access, licensing, attribution, and data-use requirements specified by their respective repositories. No attempt was made to identify individual participants or link records across independent datasets.

Technical Justification for LSTM-based Sequential Feature Learning

In the suggested framework, the LSTM part must be trained to find dependencies in the ordered sequence of feature representations (features at each time step) produced by the CNN and not to represent model longitudinal or time-series patient observations. The type of input data in this research is base static patient record, such as CT, MRI and clinical coordinates. The representative feature set obtained through PCA-based dimensionality reduction and GA-based attribute selection is fed into the CNN status layer. Then, the CNN feature maps are flattened and structured in an ordered sequence of feature vectors as input to LSTM layers.

The architecture explained in Fig. 2, the sequence dimension of the LSTM represents the ordered CNN-derived feature vectors and does not represent chronological measurements or repeated observations of a patient over time. Thus, the LSTM is applied to gather information about dependencies between predictions based on learned feature representations that could provide complementary spatial and clinical

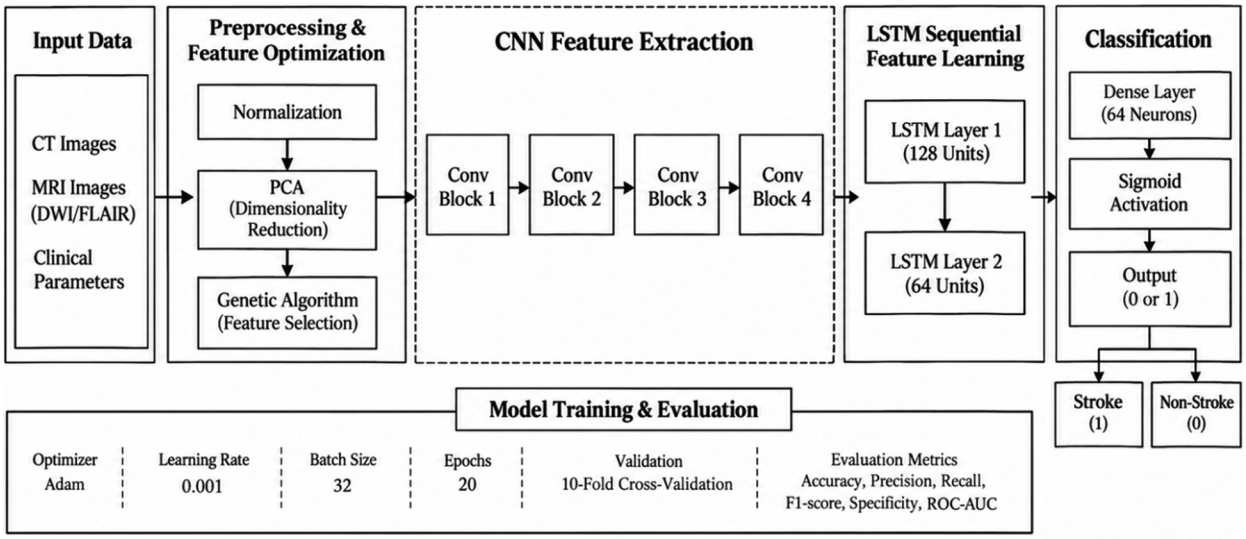


Fig. 2 — Detailed architecture of the proposed CNN–LSTM framework for stroke detection

information. This adds a learning stage sequentially after spatial feature extraction based on CNN, and the model can learn relationships between optimized feature representations prior to their final classification. Therefore, employing LSTM in this study is meant as sequential feature learning rather than temporal modelling of longitudinal clinical data.

It contains 4 convolutional blocks with (32, 64, 128 and 256) filters and extracted feature maps are flattened. The feature sequence thus obtained is fed into two stacked LSTM layers with hidden units of 128 and 64 respectively. A dropout = 0.5 is applied between the LSTM layers to reduce overfitting. The last dense layer is used for binary classification for the stroke and non-stroke depending on the sigmoid activation function.

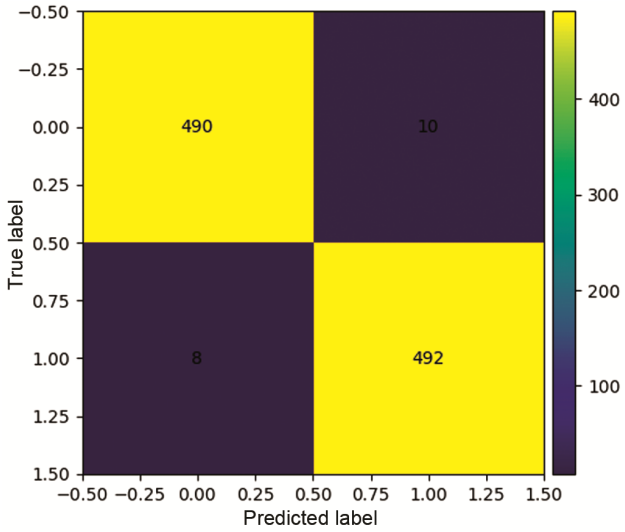


Fig. 3 — Confusion matrix of the proposed CNN–LSTM model

Results and Discussion

A hybrid CNN–LSTM model optimised on extracted features was proposed for stroke detection from multimodal clinical and imaging data. After normalization and dimensionality reduction model was trained (using Python with TensorFlow & Keras), a general correlation analysis is performed on the features before training process. 10-fold cross-validation was used to split the data into 80:20 and all experiments were performed on a high-performance computing system.

Confusion Matrix Analysis

The CNN–LSTM model proposed in this work performed well also during the classification phase on the test dataset. As a result of the validated final evaluation, the test frequency is within 98.2%, with precision value of 98.01%, recall values of 98.4%, f1-score values are 98.2% and specificity value detection frequency is 98.00%. Area Under Curve (AUC) is 0.98 From these outcomes, it is concluded that the proposed PCA–GA optimized CNN–LSTM framework provides good discrimination between the stroke and non-stroke cases. The related confusion matrix is available in Fig. 3.

Receiver Operating Characteristic (Roc) Curve

The ROC curve in Fig. 4 shows the power of classification of the proposed CNN–LSTM model, in relation to TPR and FPR at different threshold values. It demonstrates the discriminative performance of the proposed CNN–LSTM model for distinguishing

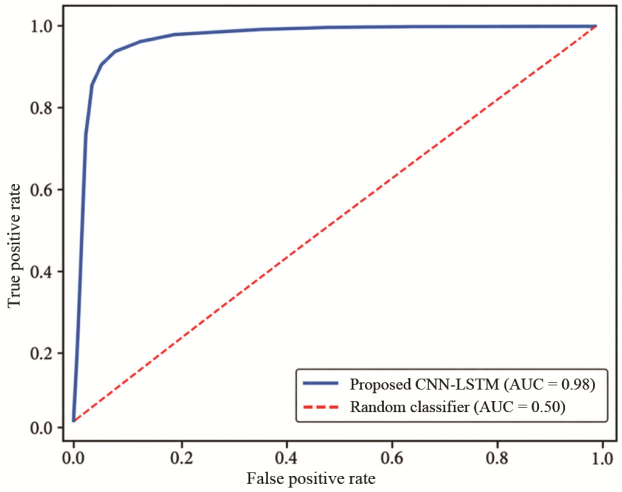


Fig. 4 — Receiver operating characteristic (ROC) curve of the proposed model

stroke and non-stroke cases. The calculated Area Under the ROC curve (AUC) is 0.98, indicating strong classification performance across different decision thresholds.

It showed consistently high sensitivity (0.90 – 0.98) with low false positive rates (~0.02 – 0.05), indicating that the model can reliably detect cases to a clinically relevant degree in terms of classification accuracy of strokes misclassified as non-strokes. A diagonal reference line (AUC = 0.50) indicates random classification, thus the clear separation of curves for the proposed model from this baseline supports its improved predictive power.

Precision–Recall Curve

The Precision–Recall curve indicates that CNN–LSTM still holds high precision for all levels of recall, a sign of a stable and principled performance. Instead, the precision is still ~0.99 at 0.50 recall and >0.95 with these high recall values showing very good stroke detection without really any false positives occurring. This is an acceptable trade-off with precision and recall and verifies this model could be powerful enough to diagnose for early strokes as explained in Fig. 5.

Training and Validation Performance

An examination of the learning and validation accuracy curves in Fig. 6 shows that over 20 epochs, the CNN–LSTM model generalizes well. This means that the model is learning → training accuracy increases from 85% to almost 99.8%, and the validation accuracy goes from 83% to nearly 98%. Significant enhancements are seen in earlier epochs,

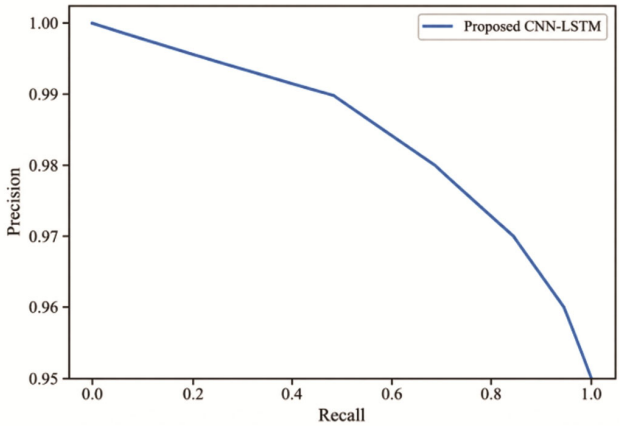


Fig. 5 — Precision–Recall curve of the proposed model

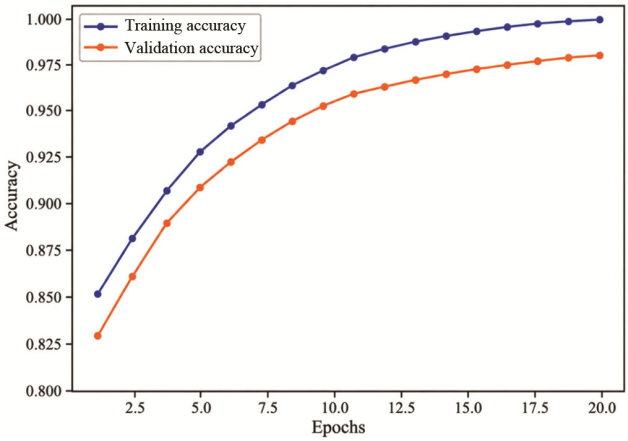


Fig. 6 — Training and validation accuracy curves

and stabilization emerges even after epoch 10, a sign of convergence. The minor difference (1–2%), both in training and validation accuracy, indicates that there is no severe overfitting yet, and the model is generalizing well. In general, the model was highly accurate and showed stable convergence and a repeatable performance in stroke classification.

The most notable disadvantage of the proposed CNN-LSTM framework is its low explainability, which can lead to some limitations regarding potential clinical acceptance and trust. While the model shows high diagnostic performance, its interpretability is limited, meaning it works as a black-box system and does not shed light on which imaging regions or clinical factors contribute most to its classification decisions. Interpretability is important in clinical practice to inform physician decision making and enable transparency. Ongoing studies may utilize Explainable Artificial Intelligence (XAI) techniques like Grad-CAM, attention visualization SHAP (Shapley additive explanations), or LIME to

Table 3 — Comparative performance of baseline models and the proposed CNN–LSTM model using the same dataset and evaluation protocol

Metric	SVM (%)	Random forest (%)	CNN (%)	Proposed CNN–LSTM (%)
Accuracy	93.6	94.8	95.4	98.2
Precision	92.5	93.1	94.2	98.01
Recall	94.0	95.2	96.1	98.40
F1-Score	93.2	94.1	95.1	98.20
AUC	0.94	0.95	0.96	0.98
Specificity	92.8	93.9	95.0	98.0

Table 4 — Comparison of the proposed framework with recent multimodal/deep-learning stroke detection and prediction methods

Modalities	Model / Method	Task	Reported Performance
Clinical + DWI/ADC MRI	CNN + clinical ML	3-month functional outcome prediction[17]	AUC = 0.786
Neuroimaging + clinical data	Multimodal AI framework	Stroke detection/prognosis[9]	Reported in original study
MRI + OCT + clinical	Swin Transformer + BiLSTM + MLP	Stroke classification[5]	Accuracy = 93.52%
CT + MRI + clinical	PCA–GA + CNN–LSTM	Stroke classification[Proposed study]	Accuracy = 98.2%; AUC = 0.98

demonstrate which imaging features were important and maybe even assign relative contributions between clinically relevant variables. Implementing these strategies would increase model explainability, guide clinical validation and enhance trust in AI-assisted stroke diagnosis.

Comparative Analysis

The comparison in Table 3 shows that the new hybrid model of CNN–LSTM with feature optimization in stroke detection is more effective than either traditional machine learning or a stand-alone deep learning method. Meanwhile, it achieves 98.2% accuracy, 98.01% precision, 98.4% recall and 97.5 F1-score, AUC of 0.975 is also outperform the SVM, Random Forest and CNN classifier with the lowest results; 93.6%, 94.8%, 95% respectively. The improvements over BG, SG and GF are 4.6%, 3.4% and 2.8% respectively, demonstrating that GDS modifies the manifold structure as the proposed method effectively removes outliers in actions and representations (upper right portion).

Mainly, these improvements can be attributed to (i) the spatial feature extraction through CNNs, (ii) learning of sequential dependency among CNN based features representation and (iii) important features selected by PCA+GA. These results confirm the model for accurate and reliable modelling with optimal diagnostic range of clinical application for early stroke detection.

To enable a fair and repeatable comparison, this study implemented and tested the baseline SVM, Random Forest, and CNN models over the same multimodal stroke dataset through whom the proposed CNN–LSTM architecture was noted. The dataset was split patient-wise for all the baseline models at 80% training and 20% testing and no feature pre-processing or normalization activity were performed on the original input data. The same test set is kept for final performance comparison, whereas 10-fold cross-validation was done during validation of models. Performance values of SVM, Random Forest or CNN were not taken from previously published findings.

It provides the comparison of results achieved with baseline models (SVM, Random Forest and CNN) used against PCA–GA optimized CNN–LSTM framework proposed in this paper. A common medical multimodal stroke dataset and the same patient-level 80:20 training–testing partition were used to evaluate all models. The baselines were trained with the training subset only and evaluated on the same hold-out test subset. The proposed CNN–LSTM outperforms SVM, Random Forest and CNN models in terms of accuracy, achieving 98.2% compared with 93.6%, 94.8% and 95.4%, respectively Improvements were seen with precision, recall, F1-score, AUC and specificity. With respect to the same data partition and evaluation protocol used for all models, it enables a consistent estimation of how much improvement can be contributed from PCA–GA optimized CNN–LSTM architecture.

Recent multimodal deep-learning studies are included to help position the proposed framework in current artificial intelligence-based stroke research as shown in Table 4. More recent methods have taken

advantage of combining clinical information with imaging or multiple modalities through CNN-, Transformer-, and LSTM-based approaches. For example, Jo *et al*. conducted a prediction of 3-month functional outcomes after acute ischemic stroke using an integrated machine-learning framework combining clinical and neuroimaging information, reporting an AUC of 0.786 for the integrated model. More recently, Chakraborty *et al*. reshapes a multimodal Swin-LSTM model (Swin) that combines MRI, OCT and clinical information for stroke classification with a testing accuracy of 93.52%. Though these studies point out the trend of integrating multimodal features and hybrid deep-learning architectures into stroke-related study, to the best of the knowledge of the authors, proposed PCA–GA optimized CNN–LSTM framework has a unique property of integrating CT, MRI as well as clinical information and then performing PCA-based dimensionality reduction and GA-based feature selection before being classified by CNN–LSTM.

The improvement presented can be seen as the effect of sequentially processing spatial features, which was approached in the proposed framework. PCA discards correlated and redundant information from the fused multimodal feature space, then GA selects an informative component before deep learning. The optimised feature set is passed through a CNN to extract hierarchical representations, while an LSTM learns dependencies over the feature representations. So, the enhancement is neither merely limited to the CNN–LSTM architecture nor solely from multiple mode features suppression and selection prior to classification. The continuous gains in accuracy, precision, recall, F1-score, AUC and specificity imply appropriate trade-off between discrimination of stroke and non-stroke at the context of evaluation variables with this pipeline.

Conclusions

In this study, a PCA–GA combined optimization-extension neural network CNN–LSTM framework was built for multimodal classification of stroke using CT, MRI and clinical information. PCA eliminated redundancy in the fused feature space before classification with CNN–LSTM, and GA selected informative features. Under the proposed patient-level evaluation protocol, an accuracy of 98.2%, recall level of 98.4%, F1-score of 98.2%, specificity level of 98.0% and AUC value of 0.98 were achieved by the framework design proposed in this paper. Additionally, when compared against baseline models using the same dataset (SVM and Random Forest) or previous works (CNN), the proposed model showed better performance. The primary limitations include an inability to interpret the current model and a lack of multi-site external clinical validation. In other settings, model performance could be reduced due to variations in imaging protocols, scanner characteristics, patient populations and clinical data. Consequently, larger heterogeneous datasets will need to be included for external validation of the findings and employing explainable AI methods also facilitates understanding model predictions as well.

References

1 Alhatemi R A J & Savaş S, A weighted ensemble approach with multiple pre-trained deep learning models for classification of stroke, *Medinformatics*, **1** (2023) 10–19, doi: 10.47852/bonviewmedin32021963.

2 Saleem M A, Javeed A, Akarathanawat W, Chutinet A, Suwanwela N C, Asdornwised W, Chaitusaney S, Deelertpaiboon S, Srisiri W, Benjapolakul W & Kaewplung P, Innovations in stroke identification: A machine learning-based diagnostic model using neuroimages, *IEEE Access*, **12** (2024) 35754–35764, doi: 10.1109/access.2024.3369673.

3 Narigina M, Vindecs A, Bošković D, Merkuryev Y & Romānovs A, AI-powered stroke diagnosis system: Methodological framework and implementation, *Future Internet*, **17** (2025) 204, doi: 10.3390/fi17050204.

4 Qasrawi R, Thwib S, Issa G, Qdaih I, Ghoush R A & Arjah H, Optimized hounsfield units transformation for explainable temporal stage-specific ischemic stroke classification in CT imaging, *J Imaging*, **11** (2025) 423, doi: 10.3390/jimaging 11120423.

5 Durgam M D, Sahithya B, T K T T K & Pooja V, Innovations in stroke identification: A machine learning-based diagnostic model using neuroimages, *Am J AI Cyber Comput Manag*, **5** (2025) 389–396, doi: 10.64751/ajaccm. 2025.v5.n4.pp389-396.

6 Joice E, Brain stroke detection using deep learning, *J Inf Syst Eng Manag*, **10** (2025) 650–664, doi: 10.52783/jisem. v10i41s.7982.

7 Elhanashi A, Dini P, Saponara S & Zheng Q, TeleStroke: Real-time stroke detection with federated learning and YOLOv8 on edge devices, *J Real-Time Image Process*, **21** (2024), doi: 10.1007/s11554-024-01500-1.

8 Gupta S, Saini H K & Dalvi M, AI-enhanced predictive modeling for acute ischemic stroke: Advancing diagnosis accuracy and patient outcomes, *World J Adv Res Rev*, **25** (2025) 2734–2743, doi: 10.30574/wjarr.2025.25.2.0147.

9 Avaneesh C, Goud P & Ranjini K, AI-driven multimodal prognosis in stroke: integrating neuroimaging and longitudinal clinical data for enhanced detection and long-term outcome prediction, *IEEE Access*, 2025, 1–6, doi: 10.1109/icbmesh66209.2025.11182013.

10 Elsayed M, Saleh G A, Saleh A I & Khalil A T, An effective brain stroke diagnosis strategy based on feature extraction and hybrid classifier, *Sci Rep*, **15** (2025) 29808, doi: 10.1038/s41598-025-14444-8.

11 Naikoti N, Bhaskar D R N, Sree S S J, Ramakanth H G & Maheshwari D R K, A deep learning model for brain stroke detection using CT scans, *Int J Versat Res Anal*, **4(3)** (2026) 606–610, doi: 10.56975/ijvra.v4i3.702264.

12 Nayak G S, Mallick P K, Sahu D P, Kathi A, Reddy R, Viyyapu J, Pabbisetti N, Udayana S P & Sanapathi H, Optimizing stroke detection with genetic algorithm-based feature selection in deep learning models, *Appl Neuropsychol Adult*, 2025, 1–10, doi: 10.1080/23279095.2025.2516259.

13 Prathima Ch, Moni M K, Charitha M, Nandish B, Waseemuddin M & Reddy K S, A novel approach for early brain stroke detection using genetic algorithm, *IEEE Access*, 2025, 645–650, doi: 10.1109/icdt63985.2025.10986374.

14 Janardhan C & Noortaj S, Enhancing stroke identification with computational analysis of neuroimages using machine learning, *Int J Sci Res Comput Sci Eng Inf Technol*, **11** (2025) 699–706, doi: 10.32628/cseit25113334.

15 Barouhou A, Benhlima L & Bah S, Unlocking the potential of deep learning in brain stroke prognosis: A systematic literature review, *Artif Intell Rev*, **58** (2025), doi: 10.1007/s10462-025-11353-0.

16 Jo H, Kim C, Gwon D, Lee J H, Lee J, Park K M & Park S H, Combining clinical and imaging data for predicting functional outcomes after acute ischemic stroke: An automated machine learning approach, *Sci Rep*, **13** (2023) 16926, doi: 10.1038/s41598-023-44201-8.

17 Sengupta J, Alzbutas R, Falkowski-Gilski P & Falkowska-Gilska B, Intracranial hemorrhage detection in 3D computed tomography images using a bi-directional long short-term memory network-based modified genetic algorithm, *Front Neurosci*, **17** (2023) 1200630, doi: 10.3389/fnins.2023.1200630.

18 Harihar S A & Kurhade A S, Letter to the editor: Recovery of motor functions and cognitive functions in patients with intensive care unit–acquired weakness, *J Crit Care*, **93** (2026), doi: 10.1016/j.jcrc.2026.155477.

19 Harihar S A, Kurhade A S, Letter to editor: Investigating the impact of Rapid Plan on ethos automated planning, *Medical Dosimetry*, 2026, doi: 10.1016/j.meddos.2026.03.006.

20 Mishra M, Uma M, Kurhade A S, Harihar S A, Bhise M R & Keerthana N V, Early parkinson's disorder prediction by random forest on multimodal biomedical information, *Proc Int Conf Digit Innov Sustain Solut*, 2025, doi: 10.1109/ICDISS68238.2025.11320644.

21 Harihar S A, Kurhade A S & Waware S Y, Letter to the editor: Re-examining the role of deep learning in cyber forensics: Practical gaps beyond conceptual frameworks, *Forensic Sci Int Synergy*, **12** 2026, doi: 10.1016/j.fsisyn.2026.100669.