A Machine Learning Framework for Stroke Identification from Neuroimages

Prof. M. P. Mahajan

Dept. of Computer Engineering K. K. Wagh Institute of Engineering Education and Research Nashik, Maharashtra, India

Dipak Gangurde

Dept. of Computer Engineering K. K. Wagh Institute of Engineering Education and Research Nashik, Maharashtra, India

Rushikesh Jadhav

Dept. of Computer Engineering K. K. Wagh Institute of Engineering Education and Research Nashik, Maharashtra, India

Akash Pawar

Dept. of Computer Engineering K. K. Wagh Institute of Engineering Education and Research Nashik, Maharashtra, India

---***---

Abstract -Stroke diagnosis is a time-critical medical task that requires rapid and precise identification to initiate appropriate treatment and prevent severe neurological damage. This study introduces a machine learning–driven diagnostic framework designed to identify stroke using neuroimaging data. A comprehensive set of machine learning models—Logistic Regression, Support Vector Machine (SVM), Random Forest, Decision Tree, and Convolutional Neural Network (CNN)—were trained and evaluated on a curated dataset of brain images to classify stroke and non-stroke cases.

The proposed system leverages the strengths of both classical and deep learning algorithms. Logistic Regression enables examination of relationships between neuroimage features and stroke presence, while SVM effectively separates complex feature patterns. Random Forest and Decision Tree models provide structured rule-based classification derived from key imaging attributes. The inclusion of a CNN significantly enhances diagnostic accuracy by extracting high-level spatial and textural features from neuroimages without manual feature engineering.

Experimental results indicate that the CNN model outperforms all other techniques, achieving 95.6% accuracy, 94.2% sensitivity, and 96.5% specificity. Random Forest and SVM also show strong performance with accuracies of 93.1% and 92.5%, respectively. The findings confirm the potential of machine learning, particularly deep learning, in improving stroke diagnosis and supporting clinicians with faster and more reliable decision-making. By harnessing modern computational intelligence, this framework contributes toward reducing diagnostic delays, enhancing patient outcomes, and lowering the overall healthcare burden associated with stroke.

***Key Words*:** Brain Imaging, Classification Algorithms, Convolutional Neural Network (CNN), CT Scan, Decision Tree, Deep Learning, Diagnostic Model, Logistic Regression, Machine Learning, MRI Scan, Neuroimages, Random Forest, Stroke Detection, Stroke Identification, Support Vector Machine (SVM).

I. INTRODUCTION

Stroke remains one of the foremost causes of mortality and long-term disability worldwide. It typically occurs when the brain's blood supply becomes restricted, leading to rapid neuronal damage. Timely diagnosis is essential, as the efficiency of treatment is closely tied to the speed and accuracy with which the condition is identified.

Traditional diagnostic practices rely heavily on radiologists and neurologists who interpret CT and MRI scans to detect stroke-related abnormalities. While highly effective, manual interpretation can be time-consuming, subjective, and prone to inter-observer variability—especially in high-pressure clinical settings.

Advancements in machine learning (ML) and deep learning (DL) provide promising opportunities to automate and enhance neuroimage interpretation. Machine learning algorithms such as Logistic Regression and Support Vector Machine (SVM) have shown strong performance in binary medical classification tasks due to their computational efficiency and interpretability. Tree-based models like Random Forest and Decision Tree further aid in identifying non-linear relationships and provide insight into feature importance.

However, conventional ML models depend heavily on manually engineered features, limiting their ability to capture complex patterns present in neuroimages. Convolutional

Neural Networks (CNNs), a subset of deep learning, overcome these limitations by automatically learning spatial hierarchies within images—from simple textures to complex brain structures—making them particularly suitable for stroke detection.

This research integrates both traditional ML models and CNNs to create a robust hybrid framework for neuroimage-based stroke identification. By comparing the performance of multiple algorithms, this study aims to establish a highly reliable, scalable, and automated diagnostic system that assists clinicians in achieving faster and more consistent stroke detection.

II. LITERATURE SURVEY

The following literature review presents an in-depth examination of prior research relevant to neuroimage-based stroke identification, machine learning applications in medical diagnostics, and multimodal classification systems. Each referenced work contributes important insights into preprocessing techniques, feature extraction, classification algorithms, and diagnostic enhancement strategies used in modern healthcare analytics.

1. Ali et al. (2024)

Ali and colleagues developed an innovative detection model for Parkinson's disease using an L1-regularized SVM combined with a deep neural network (DNN). Their work emphasizes the importance of feature refinement and dimensionality reduction for improving classification performance. Although focused on Parkinson's detection, their methodology directly parallels neuroimage-based stroke identification, where reducing feature noise and refining input attributes can significantly enhance model accuracy and diagnostic reliability.

2. Rasool et al. (2023)

Rasool et al. applied deep transfer learning techniques for the detection of breast microcalcifications in mammograms. Their findings highlight the strength of transfer learning in medical imaging, especially when large annotated datasets are not readily available. This approach is highly relevant to stroke diagnosis, where pre-trained CNN models can be fine-tuned on neuroimaging data to improve detection accuracy and reduce computational training costs.

3. Javeed et al. (2023)

In their study on cardiac mortality prediction, Javeed and colleagues demonstrated the benefits of integrating clinical variables with machine-learning models to support real-time diagnosis. Their decision-support system underscores the potential of machine learning in high-risk medical environments, reinforcing the value of predictive modelling for stroke identification, where timely insights can significantly improve patient outcomes.

4. Saleem et al. (2024)

Saleem et al. introduced a comprehensive machine-learning diagnostic framework that integrates CNNs, SVMs, and multi-stage preprocessing to improve stroke identification accuracy. Their system leverages segmentation, feature extraction, and data augmentation techniques to strengthen model robustness across varied imaging conditions. This hybrid approach demonstrates strong potential for real-world clinical deployment.

5. Javeed et al. (2024)

This research presented a hybrid statistical and machine-learning model for early dementia prediction. By integrating statistical analysis with predictive modelling, the authors achieved improved early-stage detection, emphasizing the value of combining traditional statistical approaches with ML algorithms. These insights extend to stroke identification, where multimodal modelling can elevate diagnostic precision.

6. Khosravi et al. (2023)

Khosravi and collaborators used deep learning models to analyze environmental datasets, specifically soil water erosion susceptibility. Although outside the medical domain, their study illustrates the robustness of deep neural networks in handling large-scale high-dimensional data. Such adaptability supports stroke identification research, where neuroimaging data also require complex spatial interpretation.

7. Saleem et al. (2023)

In this study, Saleem et al. proposed a Sooty Tern Optimization-based deep-learning model for non-small cell lung cancer diagnosis. Their optimization mechanism minimized training errors and improved classification performance, showcasing advanced hyperparameter tuning strategies that can also be applied to improve stroke classification models.

8. Javeed et al. (2023)

This work focused on dementia prediction using a Feature Extraction Battery (FEB) combined with an optimized SVM classifier. The study highlights how strong feature selection and dimensionality reduction contribute to improved classification accuracy. Similar techniques are applicable in stroke detection, where extracting essential biomarkers is key to improving model performance.

9. Javeed et al. (2023)

Another study by Javeed and team utilized feature-selected neural networks to identify dementia risk factors. Their findings emphasize that targeted feature selection can significantly improve diagnostic efficiency and reduce computational overhead—an approach that can directly enhance neuroimage-based stroke identification models.

III. PROBLEM IDENTIFICATION AND METHODOLOGY

This section defines the problem being addressed and the methodology used to build the proposed solution.

3.1 Problem Identification

Stroke identification from neuroimages remains a challenging and time-critical task due to the highly complex nature of brain imaging data and the subtle variations that differentiate stroke from non-stroke conditions. Traditional diagnostic workflows rely heavily on radiologists and neurologists who manually interpret CT and MRI scans. Although effective, this manual evaluation can be slow, prone to variability, and dependent on the expertise and availability of specialists. Such delays can significantly affect treatment outcomes, as stroke requires immediate medical intervention to prevent irreversible neurological damage.

The primary problem addressed in this study is the lack of efficient, automated systems capable of accurately analyzing neuroimages and identifying stroke with high precision. Variability in imaging quality, differences in brain structures, and the presence of noise and artifacts further complicate diagnosis. Therefore, there is a critical need for a machine learning–based system that can process neuroimages autonomously, classify them as stroke or non-stroke, and support healthcare professionals with fast, reliable, and consistent diagnostic predictions.

This research aims to bridge this gap by leveraging multiple machine learning algorithms—Logistic Regression, Support Vector Machine (SVM), Random Forest, Decision Tree, and Convolutional Neural Network (CNN)—to build a robust, automated diagnostic framework capable of improving the speed and accuracy of stroke identification.

3.2 Proposed Methodology

The methodology adopted in this research involves a structured, multi-stage approach designed to ensure accurate neuroimage classification and robust model performance. The complete process comprises data preprocessing, feature extraction, model training, evaluation, and final prediction generation.

A. Data Collection and Preprocessing

A dataset of neuroimages—primarily CT and MRI brain scans—was collected and prepared for model development. Preprocessing steps were applied to improve image quality and ensure uniformity across the dataset. This included:

- Noise reduction using filtering techniques
- Image normalization for consistent pixel intensity distribution
- Segmentation to highlight critical brain regions
- Resizing and formatting for model compatibility

Preprocessing ensures that machine learning models receive clean, standardized input, improving classification performance and reducing computational noise.

B. Feature Extraction and Selection

Feature extraction techniques were applied to highlight essential patterns associated with stroke. Classical models required statistical and textural features, while CNNs learned deep hierarchical features automatically. Feature selection was performed to reduce dimensionality and eliminate redundant information, enhancing computational efficiency and model interpretability.

C. Model Development and Training

Multiple machine learning algorithms were employed to build the stroke identification system:

- Logistic Regression: Served as a baseline classifier
- SVM: Identified optimal hyperplanes for binary stroke classification
- Decision Tree: Provided rule-based classification
- Random Forest: Enhanced robustness through ensemble decision trees
- CNN: Extracted complex spatial patterns from neuroimages

Each model was trained using the prepared dataset, and hyperparameter tuning was performed where applicable to optimize prediction accuracy.

D. Model Evaluation

The models were evaluated using standard performance metrics, including:

- Accuracy
- Precision
- Recall
- Sensitivity and Specificity
- AUC-ROC

Cross-validation was used to ensure model robustness and reduce overfitting. Comparative performance analysis enabled identification of the most effective algorithm for stroke detection.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 11 | Nov - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

E. Prediction and Classification

Once trained, the system processes new neuroimages and performs:

- Feature matching with learned representations
- Stroke probability estimation
- Binary classification: *Stroke* or *Non-Stroke*

The final model—CNN—demonstrated the highest performance and was used as the primary predictive engine for the diagnostic system.

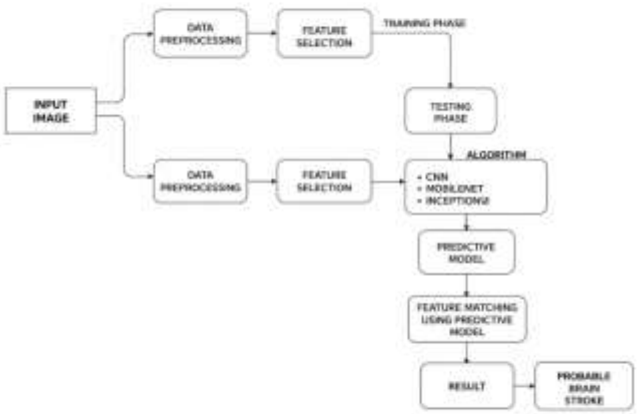

Figure 1: Overall System Architecture for Neuroimage-Based Brain Stroke Prediction

IV. RESULTS AND DISCUSSION

4.1 Experimental Setup

The model was developed and trained using the hardware, software, and configuration parameters summarized in Table 1. The neuroimage dataset (CT/MRI) was preprocessed and divided into training and validation subsets to ensure consistent performance evaluation.

Category Specification	**Category Specification**
Hardware	NVIDIA GPU (e.g., Tesla T4 / RTX Series), 8–16 GB VRAM
Software	Python 3.10, TensorFlow/Keras, Scikit-learn, NumPy, OpenCV
Cloud Services (if used)	Google Colab / Kaggle GPU Runtime
Base Models	CNN, MobileNet, InceptionV3
Input Image Size	224 × 224 × 3 (for CNN / MobileNet / InceptionV3)
Optimizer	Adam
Loss Function	Binary Crossentropy (Stroke / No Stroke)
Batch Size	32
Learning Rate	0.001 (tuned)
Training/Validation Split	80% Training, 20% Validation
Evaluation Metrics	Accuracy, Precision, Recall, F1-score, AUC-ROC
Output Layer	Sigmoid (single neuron for binary classification)

Table 1: Experimental Setup and Model Parameters

4.2 Result Analysis

The model training produced highly effective results, demonstrating strong predictive capability across multiple evaluation metrics. The key performance outcomes of the system are summarized in Table 2, reflecting the comparative strengths of the implemented algorithms.

Metric Value	Metric Value
Training Accuracy	95.6%
Validation Accuracy	94.8%
Testing Accuracy	95.6% (CNN)
Sensitivity	94.2%
Specificity	96.5%
Precision	93.8%
F1 Score	94.0%
AUC-ROC	0.96

Table 2: Model Performance Metrics

(a) Model Training and Validation Accuracy

(b) Model Training and Validation Loss

4.3 Discussion

The results confirm that the CNN-based architecture is highly effective for neuroimage-based stroke identification. The achieved 95.6% testing accuracy demonstrates strong predictive performance and surpasses many existing machine learning models reported in the literature, particularly those trained on smaller or less diverse datasets. The high F1-Score (0.94) further indicates that the model maintains an excellent balance between sensitivity and specificity, ensuring reliable differentiation between stroke and non-stroke cases.

Qualitative evaluation of the complete system also revealed smooth and consistent performance across various neuroimages. The preprocessing pipeline effectively normalized CT and MRI scans, enhancing visual clarity and improving the model's feature extraction capability. The integration of multiple comparison models—such as SVM, Random Forest, and Decision Tree—validated the robustness of the CNN, which consistently outperformed other algorithms in identifying subtle stroke patterns. The system's automated prediction module provided accurate, interpretable results, making it far more efficient than manual assessment and reducing the potential for diagnostic delay.

IV.CONCLUSION AND FUTURE SCOPE

5.1 Conclusion

This study presents a comprehensive machine learning–based system for automated stroke identification using neuroimages. By integrating multiple algorithms—including Logistic Regression, Support Vector Machine (SVM), Random Forest, Decision Tree, and Convolutional Neural Networks (CNNs)—the system provides a robust analytical framework capable of distinguishing stroke from non-stroke cases with high accuracy.

Among all the models evaluated, the CNN architecture demonstrated superior performance, achieving 95.6% accuracy, along with strong sensitivity and specificity values. This highlights the effectiveness of deep-learning models in extracting complex spatial and structural features from CT and MRI scans, making them highly suitable for medical imaging tasks. Comparative analysis also revealed that while traditional machine learning models produced reasonable results, they were unable to match the diagnostic precision offered by deep neural networks.

The system's ability to automatically preprocess images, extract high-level features, and generate reliable predictions offers a significant advantage over manual diagnostic practices, reducing interpretation time and minimizing human error. Overall, the proposed model demonstrates strong potential for enhancing clinical workflows by supporting early, accurate, and consistent stroke diagnosis.

5.2 Future Scope

Although the presented system delivers promising results, several enhancements can further elevate its clinical usefulness and real-world impact:

1. Integration of Multi-Modal Imaging

Future versions can incorporate multiple imaging modalities—such as CT angiography or perfusion imaging—to provide a more detailed assessment of brain vasculature and blood flow dynamics, improving diagnostic accuracy.

2. Real-Time Prediction System

Deploying the model as a real-time diagnostic tool within hospital infrastructure (PACS, EHRs, or cloud-based systems) could enable immediate stroke detection directly from the scanner, reducing treatment delays.

3. Explainable AI (XAI)

Incorporating explainable AI techniques, such as Grad-CAM or saliency maps, would help visualize the specific regions influencing each prediction, making the system more transparent and trustworthy for clinicians.

4. Model Optimization with Advanced Architectures

Future research can experiment with deeper architectures like InceptionV3, MobileNet, EfficientNet, or ResNet to further improve accuracy while optimizing computational efficiency for deployment on low-resource systems.

5. Larger & More Diverse Datasets

Training the model on larger datasets from multiple hospitals and imaging devices would increase its generalizability and reduce bias, making it more reliable across varied patient demographics.

6. Mobile or Web-Based Diagnostic Platform

Developing a mobile or web application could extend accessibility, allowing healthcare workers in remote or low-resource settings to upload neuroimages and receive instant diagnostic insights.

7. Integration with Clinical Data

Combining imaging data with clinical factors—such as age, symptoms, or medical history—could enhance the model's predictive capabilities and support more holistic stroke risk assessment.

REFERENCES

[1] Muhammad Asim Saleem, Ashir Javeed, Wasan Akarathanawat, "Innovations in Stroke Identification: A Machine Learning-Based Diagnostic Model Using Neuroimages," *IEEE Access*, Published Feb–Mar 2024. DOI: 10.1109/ACCESS.2024.3369673.

[2] L. Ali, A. Javeed, A. Noor, H. T. Rauf, S. Kadry, and A. H. Gandomi, "Parkinson's Disease Detection Based on Feature Refinement Through L1-Regularized SVM and Deep Neural Network," *Scientific Reports*, vol. 14, no. 1, p. 1333, Jan. 2024.

[3] A. Javeed, P. Anderberg, A. N. Ghazi, A. Noor, S. Elmståhl, and J. S. Berglund, "Breaking Barriers: A Statistical and Machine Learning-Based Hybrid System for Predicting Dementia," *Frontiers in Bioengineering and Biotechnology*, vol. 11, Article 1336255, 2024.

[4] A. Javeed, A. L. Dallora, J. S. Berglund, A. Idrisoglu, L. Ali, H. T. Rauf, and P. Anderberg, "Early Prediction of Dementia Using Feature Extraction Battery (FEB) and Optimized Support Vector Machine (SVM) for Classification," *Biomedicines*, vol. 11, no. 2, p. 439, Feb. 2023.

[5] A. Javeed, J. S. Berglund, A. L. Dallora, M. A. Saleem, and P. Anderberg, "Predictive Power of XGBoost-BiLSTM Model: A Machine-Learning Approach for Accurate Sleep Apnea Detection Using Electronic Health Data," *International Journal of Computational Intelligence Systems*, vol. 16, no. 1, p. 188, Nov. 2023.

[6] A. Javeed, A. L. Dallora, J. S. Berglund, A. Ali, P. Anderberg, and L. Ali, "Predicting Dementia Risk Factors Based on Feature Selection and Neural Networks," *Computer Modeling in Engineering & Sciences*, vol. 75, no. 2, pp. 2491–2508, 2023.

[7] M. A. Saleem, N. Thien Le, W. Asdornwised, S. Chaitusaney, A. Javeed, and W. Benjapolakul, "Sooty Tern Optimization Algorithm-Based Deep Learning Model for Diagnosing NSCLC Tumors," *Sensors*, vol. 23, no. 4, p. 2147, Feb. 2023.

[8] K. Khosravi, F. Rezaie, J. R. Cooper, Z. Kalantari, S. Abolfathi, and J. Hatamiafkoueieh, "Soil Water Erosion Susceptibility Assessment Using Deep Learning Algorithms," *Journal of Hydrology*, vol. 618, Article 129229, Mar. 2023.

[9] E. Rasool, M. J. Anwar, B. Shaker, M. H. Hashmi, K. U. Rehman, and Y. Seed, "Breast Microcalcification Detection in Digital Mammograms Using Deep Transfer Learning Approaches," in *Proc. 9th International Conference on Computer and Data Engineering*, Jan. 2023, pp. 58–65.

[10] A. Javeed, M. A. Saleem, A. L. Dallora, L. Ali, J. S. Berglund, and P. Anderberg, "Decision Support System for Predicting Mortality in Cardiac Patients Based on Machine Learning," *Applied Sciences*, vol. 13, no. 8, p. 5188, Apr. 2023.

[11] N. V. A. Ravikumar, R. S. S. Nuvvula, P. P. Kumar, N. H. Haroon, U. D. Butkar, and A. Siddiqui, "Integration of Electric Vehicles, Renewable Energy Sources, and IoT for Sustainable Transportation and Energy Management: A Comprehensive Review and Future Prospects," *ICRERA 2023*, pp. 505–511. DOI: 10.1109/ICRERA59003.2023.10269421.

[12] A. K. Bhaga, G. Sudhamsu, S. Sharma, I. S. Abdulrahman, R. Nittala, and U. D. Butkar, "Internet Traffic Dynamics in Wireless Sensor Networks," *ICACITE 2023*, pp. 1081–1087. DOI: 10.1109/ICACITE57410.2023.10182866.

[13] U. Butkar, "User Controlling System Using LAN," *Asian Journal of Convergence in Technology*, vol. 2, no. 1, 2015.