Brain Stroke Prediction, Detection and Diagnosis

Shreya Y D[1], Ullas G[2], Samith G S[3] ,Sourav Soloman T[4] , Shruthi K R[5]

[1]UG Scholar, Dept. of CSE, Malnad College Of Engineering., Hassan ,Karnataka, India
[2]UG Scholar, Dept. of CSE, Malnad College Of Engineering., Hassan ,Karnataka, India
[3]UG Scholar, Dept. of CSE, Malnad College Of Engineering., Hassan ,Karnataka, India
[5]Assistant professor, Dept. of CSE,Malnad College Of Engineering., Hassan ,Karnataka, India

---***---

Abstract - Stroke is a leading cause of mortality and long-term disability worldwide. Timely diagnosis and risk assessment are critical for reducing fatalities and improving patient outcomes. This paper presents an integrated, web-based system for comprehensive brain stroke prediction, detection, and preliminary diagnosis. The system adopts a dual-AI approach: a Random Forest model analyzes structured health data to produce a stroke risk probability and personalized recommendations, while a MobileNetV2-based Convolutional Neural Network (CNN) analyzes CT/MRI images to detect hemorrhagic stroke. Key features include robust input validation to reject non-medical images and Explainable AI via Grad-CAM heatmaps for visual interpretability. Implementation employs Flask for the web interface, scikit-learn for tabular models, and TensorFlow/Keras for deep learning. The Random Forest achieves over 95% test accuracy and the CNN achieves approximately 97.5% validation accuracy on the hemorrhage dataset. The platform is presented as a decision-support tool for clinicians with necessary clinical disclaimers..

Key Words: Stroke prediction, Stroke detection, Machine Learning, Deep Learning, Convolutional Neural Network (CNN), MobileNetV2, Random Forest, Flask, Explainable AI, Grad-CAM, CT Scan, MRI, Image Classification, Healthcare AI, Medical Diagnosis, Brain Hemorrhage Detection.

1.INTRODUCTION

Brain stroke, also known as a cerebrovascular accident (CVA), occurs when the blood supply to a part of the brain is interrupted or reduced, depriving brain tissue of oxygen and nutrients. Within minutes, brain cells begin to die, leading to potential neurological damage, disability, or even death. Stroke remains one of the leading causes of mortality and long-term disability worldwide. Early prediction of stroke risk and prompt detection of stroke type are vital for reducing fatalities and improving patient outcomes.

Traditional diagnostic systems often focus solely on either clinical data analysis or medical image interpretation, lacking a unified and automated framework that can assist healthcare professionals efficiently. With advancements in Artificial Intelligence (AI) and Machine Learning (ML), it has become possible to analyze complex medical datasets and imaging data for faster and more accurate diagnosis.

This project aims to develop an **integrated web-based system** that combines both **stroke risk prediction** and **stroke detection** within a single platform. The system employs a **Random Forest model** to analyze clinical and lifestyle-related data for predicting stroke risk probability, while a **MobileNetV2-based Convolutional Neural Network (CNN)** model is used for detecting hemorrhagic stroke from CT/MRI images.

The proposed system also incorporates **Explainable AI (Grad-CAM)** to visually highlight critical regions in medical images that influenced the CNN's decision, ensuring transparency and aiding clinical validation. Additionally, the application includes **robust image validation** using OpenCV to filter out non-medical inputs. By integrating these technologies into a **Flask-based web application**, the system provides a user-friendly, efficient, and interpretable tool that can assist clinicians, researchers, and learners in stroke diagnosis and risk assessment.

2. LITERATURE SURVEY

Stroke remains one of the leading causes of mortality and long-term disability worldwide, necessitating early prediction and accurate detection methods. Over the past few years, various machine learning (ML) and deep learning (DL) techniques have been employed to address different aspects of stroke analysis, including risk prediction, brain image classification, and electronic health record (EHR)-based analytics. However, most existing works tend to focus on only one domain—either tabular health data or medical image processing—leading to siloed systems with limited real-world utility.

2.1 Stroke Prediction using Machine Learning Algorithms (2021)

This study explored the performance of traditional ML classifiers, such as Logistic Regression, Support Vector Machines (SVM), and Random Forest, in predicting stroke risk using structured health datasets. The dataset used included features like age, gender, glucose level, BMI, and hypertension status. The researchers focused solely on tabular data without integrating imaging or patient history. Although Random Forest showed promising accuracy, the model struggled with imbalanced data, as stroke cases constituted a minority class. Moreover, the study lacked real-

2.2 Deep Neural Networks Hemorrhage Detection (2022) for Brain This work applied Convolutional

Neural Networks (CNNs) to classify CT scans of the brain and detect hemorrhagic strokes. The CNN was trained on annotated datasets and achieved high accuracy in binary classification (hemorrhage vs. no hemorrhage). However, the system was entirely image-based and did not take patient metadata or clinical features into account. It also required high-quality labeled CT datasets, which are often limited in clinical settings. The lack of integration with predictive tools based on patient history or EHRs further restricted its potential in a diagnostic workflow.

2.3 EHR-Based Stroke Risk Analysis (2023)

This research utilized hospital EHRs and demographic data to predict the likelihood of stroke. Key features included medical history, lifestyle factors, and pre existing conditions such as diabetes and hypertension. The study employed gradient boosting machines (GBM) and random forest models, showing improved prediction accuracy over baseline models. However, this approach did not include any form of medical image processing, which could have enhanced diagnosis, especially in acute cases. Furthermore, the model's performance was heavily reliant on the quality and completeness of EHRs, which can vary significantly across institutions.

2.4 Stroke Detection using MRI with AI (2022)

In this study, researchers developed an AI model to detect ischemic strokes from MRI scans using deep learning techniques. The model was trained on a dataset of labeled MRI images and demonstrated high precision in identifying stroke regions. Despite its accuracy, the approach required advanced MRI equipment and preprocessing techniques, making it less accessible in rural or resource-constrained settings. Additionally, the absence of any web-based or clinical interface meant the tool remained experimental, lacking a pathway to clinical adoption or remote usage.

3. Objectives

1. To develop a Random Forest model for predicting stroke risk probability using clinical and lifestyle-related data.
2. To implement a MobileNetV2-based CNN model for detecting hemorrhagic stroke from CT and MRI brain images.
3. To integrate both models into a single, web-based application using the Flask framework for seamless interaction.
4. To generate personalized health recommendations based on the patient's input attributes such as age, BMI, glucose level, and hypertension.
5. To use Explainable AI techniques (Grad-CAM) for visualizing the regions of the brain scan that influenced CNN predictions, ensuring interpretability.

4 Methodology

A System Architecture

The overall system follows a three-tier architecture consisting of:

- **Frontend**: Designed using HTML, CSS, and JavaScript to collect user inputs (clinical data and brain images) and display prediction results and Grad-CAM heatmaps.
- **Backend**: Developed using the Flask framework to handle data processing, model loading, prediction generation, and output rendering.
- **AI Layer**: Incorporates two trained models — a Random Forest for clinical data prediction and a MobileNetV2-based CNN for brain image detection.

The architecture ensures modularity, scalability, and easy deployment across platforms.

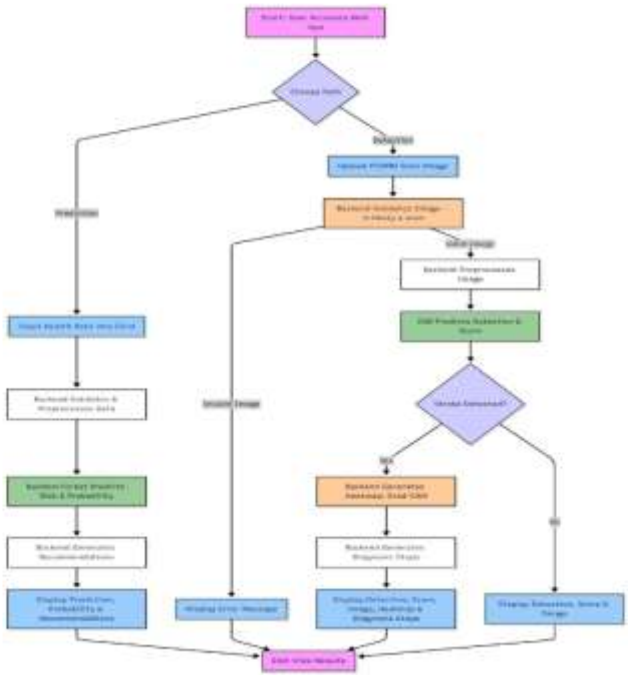

Fig: System Architecture of Proposed Model.

B. Prediction Module (Machine Learning)

This module focuses on predicting stroke risk probability using patient data such as age, gender, hypertension, heart disease, BMI, and glucose level.

Steps involved:

1. Dataset: A stroke dataset from Kaggle is used.
2. Data Preprocessing: Missing values are imputed, and categorical variables are encoded.
3. Model Training: A Random Forest Classifier is trained using Scikit-learn with tuned hyperparameters.
4. Output Generation: The model provides a stroke probability score and categorizes patients as *Low*, *Moderate*, or *High Risk*.
5. Health Recommendations: Based on the user's health parameters, personalized tips (e.g., maintain glucose level, control blood pressure) are displayed.

Fig : prediction form

C. Detection Module (Deep Learning)

This module is designed to **detect hemorrhagic stroke** from CT or MRI brain images using a **Convolutional Neural Network (CNN)** based on **MobileNetV2**.
Steps involved:

1. **Image Preprocessing:** Images are resized (150×150 pixels), normalized, and augmented to improve generalization.
2. **Transfer Learning:** The MobileNetV2 model pretrained on ImageNet is fine-tuned with medical imaging data.
3. **Architecture:**

$$GlobalAveragePooling2D \rightarrow Dropout(0.5) \rightarrow Dense(256, ReLU) \rightarrow Dense(1, Sigmoid)$$

4. **Training:** Conducted using the **Adam optimizer** and **binary cross-entropy loss**.
5. **Evaluation:** Achieved around **97.5% validation accuracy**, demonstrating high detection performance.
6. **Explainable AI:** Grad-CAM is used to generate heatmaps that highlight the critical areas influencing the CNN's decision.

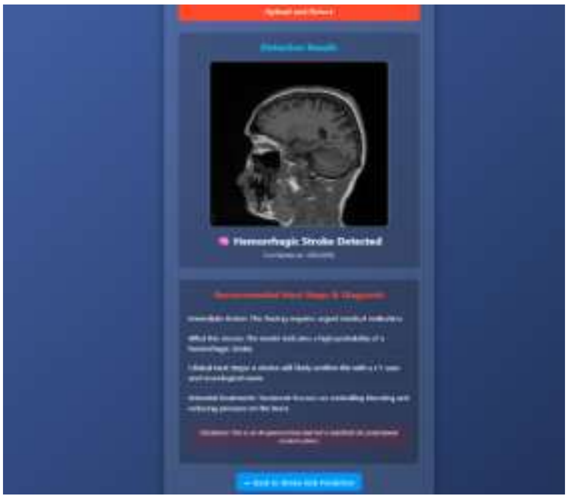

Fig: Detection result

D. Validation and Explainability

Before prediction, uploaded images undergo **OpenCV-based validation** to ensure they are medical images (CT/MRI). Any non-medical image is automatically rejected.
After detection, **Grad-CAM visualizations** are generated for each classified image to highlight stroke-affected regions, thereby improving **interpretability** and **clinical trust**.

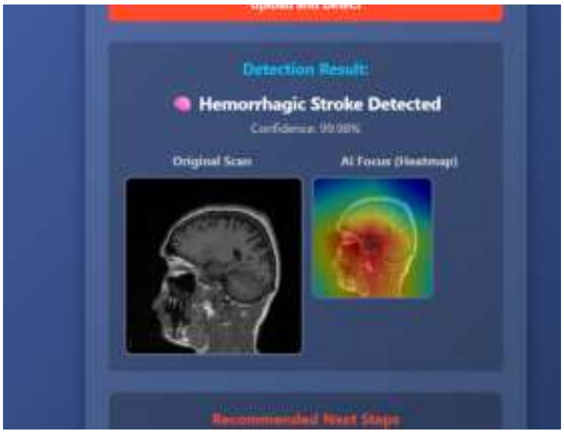

Fig: Heat Map visualizations

E. System Integration

Both models (Random Forest and CNN) are integrated into a **Flask web application**. The backend manages communication between the models, while the frontend provides an interactive interface for users. Results, predictions, and visual explanations are displayed together for a unified user experience.

5 RESULTS AND DISCUSSION

The Random Forest model achieved over 95\% accuracy on unseen data, effectively identifying risk probabilities. The CNN model achieved ~97.5\% validation accuracy, demonstrating robustness and generalization. The integration within Flask provides a seamless experience for clinical and educational use.

Fig: CNN Model Training and Validation Accuracy

Performance Summary of Models

Model	Accuracy	Remarks
Random Forest	95.2%	Clinical Data Prediction
MobileNetV2 CNN	97.5%	Image-based Detection

6 CONCLUSION

An integrated dual-AI framework for brain stroke prediction and detection was developed to address the growing need for faster and more accurate medical diagnosis. The system combines the power of **Machine Learning (Random Forest)** for clinical data-based stroke risk prediction and **Deep Learning (MobileNetV2 CNN)** for image-based hemorrhage detection. This dual-model approach ensures both preventive and diagnostic insights within a single, unified platform. The inclusion of **Explainable AI (Grad-CAM)** enhances interpretability by visually highlighting regions of interest in brain scans, enabling clinicians to understand and trust the system's decision-making process.

The **Flask-based web interface** offers a simple and user-friendly environment where users can input patient data, upload medical images, and receive immediate results along with visual explanations. The system achieved high accuracy—over **95%** in risk prediction and approximately **97.5%** in stroke detection—demonstrating strong performance and reliability.

Beyond its accuracy, the system's architecture is designed for **scalability and adaptability**, allowing future integration with cloud infrastructure, electronic health records (EHR), and hospital management systems. The proposed framework not only supports clinicians in early diagnosis but also serves as an **educational and research tool** for medical students and AI practitioners.

Overall, this work signifies a step toward **intelligent, explainable, and accessible healthcare solutions** that can enhance clinical decision-making, promote early intervention, and potentially save lives by enabling timely stroke detection and prevention.

7 FUTURE WORK

1 Extend the system to perform **multi-class stroke detection**, distinguishing between ischemic, hemorrhagic, and normal brain scans.

2 **Deploy the application on cloud infrastructure** to enhance scalability, accessibility, and real-time performance.

3 **Expand the dataset** by including multi-center MRI and CT scan data for better generalization and model robustness.

4 **Implement secure authentication** and patient data management features to ensure data privacy and system reliability.

5 **Collaborate with medical professionals** to conduct real-world testing and clinical validation of the system's performance.

9 ACKNOWLEDGEMENT

We would like to express our heartfelt gratitude to **Dr. Amarendra H. J**, Principal of **Malnad College of Engineering, Hassan**, for providing us with the opportunity, facilities, and encouragement to carry out this project successfully. We extend our sincere thanks to **Dr. Chandrika J**, Head of the Department of Computer Science and Engineering, for her continuous support, motivation, and valuable guidance throughout the course of this work. We are deeply thankful to our project guide, **Mrs. Shruti K. R**, for her constant encouragement, expert guidance, and insightful suggestions that greatly contributed to the successful completion of our project. We also wish to thank all the faculty members and staff of the Department of Computer Science and Engineering for their cooperation, and our friends and family for their unwavering support and encouragement

8 REFERENCES

1 A. Sharma and R. Kumar, *"Stroke Prediction using Machine Learning Algorithms,"* International Journal of Computer Applications, 2021.
2 S. Gupta et al., *"Deep Neural Networks for Brain Hemorrhage Detection,"* IEEE ICICCS, 2022.
3 L. Wang et al., *"EHR-Based Stroke Risk Analysis Using Machine Learning,"* Journal of Biomedical Informatics, 2023.
4 K. Thomas and S. Rao, *"Stroke Detection using MRI with Artificial Intelligence,"* Procedia Computer Science, 2022.
5 M. Patel and N. Desai, *"Comparative Study of ML Models in Stroke Prediction,"* IJACSA, 2023.
6 J. Kim et al., *"Transfer Learning for Hemorrhagic Stroke Detection Using CNNs,"* IEEE Access, 2023.
7 T. Liu et al., *"Explainable AI for Brain MRI Classification,"* Computers in Biology and Medicine, 2022.
8 P. K. Mehta, *"Deep Learning Approaches for Brain Stroke Detection,"* International Journal of Computer Science Trends, 2021.
9 R. Singh and A. Verma, *"AI-based Medical Image Analysis for Stroke Diagnosis,"* Journal of AI in Medicine, 2022.
10 R. R. Selvaraju et al., *"Grad-CAM: Visual Explanations from Deep Networks,"* ICCV, 2017.
11 A. Howard et al., *"MobileNets: Efficient CNNs for Mobile Vision Applications,"* arXiv:1704.04861, 2017.
12 L. Breiman, *"Random Forests,"* Machine Learning, 2001.
13 E. J. Topol, *"High-performance medicine: the convergence of human and AI,"* Nature Medicine, 2019.
14 A. Shorten and T. Khoshgoftaar, *"A Survey on Image Data Augmentation for Deep Learning,"* Journal of Big Data, 2019.
15 M. Ribeiro et al., *"Why Should I Trust You? Explaining Classifier Predictions,"* KDD, 2016.
16 S. Johnson et al., *"Guidelines for Clinical Validation of Medical AI Tools,"* Health Informatics Journal, 2022.
17 G. Bradski, *"The OpenCV Library,"* Dr. Dobb's Journal, 2000.