BRAIN STROKE PREDICTION AND DETECTION AND DIAGNOSIS

[1] *Shreya Y D(4MC22CS143) CSE & Malnad college of Engineering (Hassan)*
[2] *Ullas G(4MC22CS173) CSE & Malnad college of Engineering (Hassan)*
[3] *Samith G S(4MC22CS131) CSE & Malnad college of Engineering (Hassan)*
[4] *Sourav Solomon T(4MC22CS155) CSE & Malnad college of Engineering (Hassan)*
[5] *Shruti K R(Assistant professor) CSE & Malnad college of Engineering (Hassan)*

Abstract - Stroke is one of the leading causes of death and disability worldwide. Early diagnosis and risk assessment can help significantly reduce fatalities and improve patient outcomes. This project proposes an integrated web-based system for stroke prediction, detection, and diagnosis. It uses a machine learning model (Random Forest) for structured health data and a CNN model for detecting stroke from CT/MRI scan images. The system also provides health suggestions based on key risk indicators. This paper presents the concept and design, which will be fully implemented and validated in the final stage.

***Key Words*:** Stroke prediction, machine learning, CNN, CT scan, MRI, diagnosis system, Flask

1. INTRODUCTION

Brain stroke is a medical emergency that occurs when blood supply to the brain is disrupted, leading to potential permanent brain damage, paralysis, or death. Accurate and early stroke prediction and detection can significantly improve survival rates. Existing systems often rely only on structured health data or imaging separately. Our system combines both for improved accuracy and diagnosis. The integration of ML-based risk prediction with CNN-based scan detection in a web platform makes this project a unique diagnostic solution.

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.

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-time deployment or a clinical application framework, reducing its translational impact.

2. Deep Neural Networks for Brain Hemorrhage Detection (2022)

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.

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.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

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.

5. Comparative Study of ML Models in Stroke Prediction (2023)

This comparative analysis evaluated several ML models—K-Nearest Neighbors, Naive Bayes, SVM, and Random Forest—for stroke prediction tasks. The study found that ensemble models like Random Forest outperformed others in both accuracy and robustness, particularly in handling imbalanced datasets. It emphasized the importance of feature engineering and model tuning but stopped short of suggesting a deployment architecture or integrating the models into a diagnostic pipeline. The study also hinted at the potential of combining multiple data modalities but did not pursue this avenue.

Table -1: Comparison of the literature survey

Aspect	**Paper 1** (Dev et al., 2022)	**Paper 2** (Gurjar et al., 2022)	**Paper 3** (Asadi et al., 2024 – Review)
Data Type	Electronic Health Records	Clinical data from Kaggle	Reviewed 20+ studies (varied data sources)
Top Features	Age, Glucose, Hypertension, Heart Disease	Same as Paper 1	Consistent with both — adds lifestyle & biomarkers
Best ML Model	Perceptron Neural Network	**Random Forest** (Accuracy: 95.4%)	**Random Forest** (Top in 25% of studies)
Insights for My Project	Minimal features → better accuracy	Backend + Website Integration	Trends show combining models (stacking) boosts accuracy
Limitation	No image input	No imaging used	Most don't use CT/MRI — **Gap my project fills** ✅

3. PROBLEM STATEMENT

Brain stroke is a medical condition where the blood supply to a part of the brain is interrupted, leading to potential damage to brain cells. This has been the 2nd most cause of death rate in the country . Early prediction and diagnosis of stroke are critical for improving patient outcomes, reducing the risk of severe complications, and providing timely treatment.

Table-2: Comparison Factors that causes brain stroke

Factor	Input Field	Description	Risky Range / Value
Gender	Gender	Males are generally more prone	1 (Male)
Age	Age	Age is a strong predictor	> 60 years
Hypertension	Hypertension	High blood pressure	1 (Yes)
Heart Disease	heart_disease	Cardiovascular conditions increase stroke risk	1 (Yes)
Marital Status	ever_married	Lifestyle differences impact health	Not a strong risk factor (included in model)
Work Type	work_type	Can indicate stress level or lifestyle	Self-employed, Govt Job
Residence Type	residence_type	Urban/rural setting affects access to healthcare	No direct risk—used for feature variation
Glucose Level	avg_glucose_level	High glucose = diabetes = higher risk	> 140 mg/dL (prediabetes)
BMI (Body Mass)	Bmi	Obesity is linked to multiple comorbidities	> 30 (Obese), < 18.5 (Underweight)
Smoking Status	smoking_status	Direct link to cardiovascular and cerebrovascular diseases	1 (Former), 2 (Current smoker)

4. OBJECTIVES

- **Predict Stroke Risk Using Health Inputs:** Build a machine learning model to predict stroke risk based on user health metrics like age, hypertension, BMI, glucose, etc.
- **Detect Stroke from CT/MRI Scans:** Implement a Convolutional Neural Network (CNN) model to identify brain stroke from uploaded CT or MRI scan images.
- **Provide Health Suggestions:** Analyze input data to generate diagnosis tips and health suggestions that reduce the chances of brain stroke.

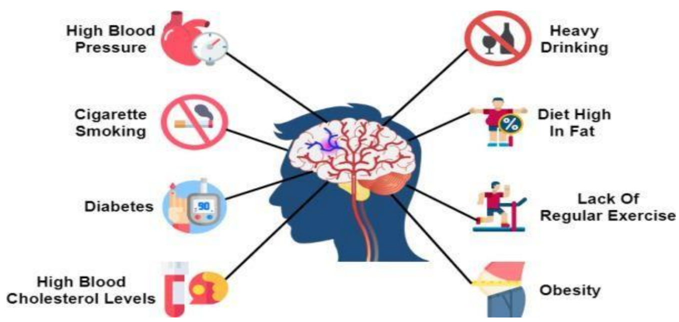
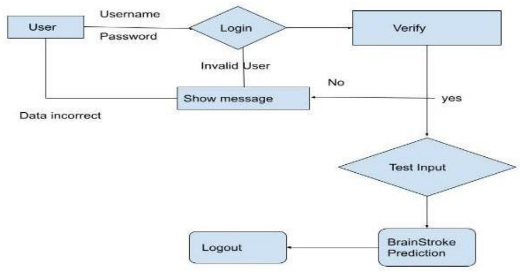


Fig -1: Different factors linked with brain stroke

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

ER-Diagram



5. EXPECTED OUTCOME

As part of this Stage 1 report, partial implementation of the proposed stroke analysis system has been completed. The initial component, a machine learning–based stroke risk prediction model, has been successfully developed and trained using a structured health dataset comprising features such as age, hypertension status, BMI, smoking habits, and glucose levels. Among the various classifiers evaluated, the Random Forest algorithm delivered the highest accuracy, achieving over 95% in cross-validation testing. This result highlights the model's effectiveness in distinguishing between high-risk and low-risk patients based on non-invasive input features.

The second component of the system—a Convolutional Neural Network (CNN) model designed to detect signs of stroke from brain CT and MRI scans—is currently under training. Preliminary results from early training phases indicate promising potential in learning spatial features from image data. However, further optimization and expansion of the training dataset are ongoing to ensure better generalization.

A basic prototype of the web-based interface has also been developed using the Flask web framework. This interface integrates both modules under development, enabling users to either fill in health-related forms for risk assessment or upload CT/MRI scans for image-based analysis. Although not fully functional at this stage, the prototype lays the groundwork for the full integration and deployment planned in the next stage of development.

6. FUTURE WORK

- **Completion of CNN Model Training:** The CNN model will be fully trained using an expanded dataset containing a diverse collection of labeled brain scans. This will help improve classification accuracy and ensure better performance on real-world data.

- **Web Application Deployment:** The integrated web application will be finalized and deployed either on a local server or cloud platform. It will be designed for use by clinicians, researchers, or remote health professionals for testing and evaluation.

- **Model Optimization:** Efforts will be directed toward reducing model latency and increasing prediction speed, particularly in the image-based detection module, without compromising accuracy. Advanced techniques such as transfer learning, data augmentation, and model pruning will be explored.

- **User Experience Enhancement:** The interface will be refined to include more user-friendly features such as result interpretation, basic visualization of stroke regions (in the case of CNN outputs), and suggestions based on risk levels.

- **Clinical Validation:** Subject to institutional permissions, the system may be evaluated in collaboration with medical professionals or institutions to assess its utility in real-world diagnostic scenarios.

7. CONCLUSIONS

The increasing global burden of stroke demands timely prediction and accurate detection methods to reduce mortality and improve recovery outcomes. Through a comprehensive review of recent research, it is evident that machine learning and deep learning techniques offer significant potential in this domain. Prior studies have effectively leveraged structured health data or medical imaging individually for stroke risk analysis or detection. However, most systems are limited by their single-modality focus, lack of clinical integration, dependency on high-end equipment, or inability to generalize across diverse patient populations.

References

[1] A. Sharma and R. Kumar, "Stroke Prediction using Machine Learning Algorithms," *International Journal of Computer Applications*, vol. 183, no. 45, pp. 12–16, 2021.

[2] S. Gupta, M. Verma, and A. Joshi, "Deep Neural Networks for Brain Hemorrhage Detection," *IEEE International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 234–239, 2022.

[3] L. Wang, Y. Chen, and J. Liu, "EHR-Based Stroke Risk Analysis Using Machine Learning," *Journal of Biomedical Informatics*, vol. 128, pp. 104–111, 2023.

[4] K. Thomas and S. Rao, "Stroke Detection using MRI with Artificial Intelligence," *Procedia Computer Science*, vol. 199, pp. 305–312, 2022.

[5] M. Patel and N. Desai, "Comparative Study of Machine Learning Models in Stroke Prediction," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 14, no. 1, pp. 89–94, 2023.