RESEARCH ARTICLE OPEN ACCESS

Brain Stroke Prediction Using Machine Learning and Power BI

A.Mary Agnalla*, Ms.P.Sathya**
*(Student ,Computer Science and Engineering, Francis Xavier Engineering College, Tirunelveli
maryagnallaa.ug.21.cs@francisxavier.ac.in)
** (Assistant Professor ,Information and Technology , Francis Xavier Engineering College, Tirunelveli
sathyap@francisxavier.ac.in)

--************************----------------------------------

Abstract:

Keywords — Brain Stroke Prediction,Machine Learning,Model Evaluation,SMOTE ,Power BI Dashboard,Accuracy,Precision,Recall,F1-Score

--************************----------------------------------

I. INTRODUCTION

In recent years, machine learning and artificial intelligence have played a transformative role in the healthcare sector, especially in disease prediction and diagnosis. Among various health conditions, stroke remains one of the leading causes of death and long-term disability worldwide. Early prediction and timely intervention are critical to reducing its impact. Leveraging machine learning for stroke prediction offers a promising approach to analyze medical data and identify high-risk individuals before the onset of a stroke. This project, titled “Brain Stroke Prediction Using Machine Learning,” focuses on building an intelligent system that can analyze patient health data to predict the likelihood of a stroke. By using a supervised machine learning approach, the system learns from historical medical records and identifies patterns and correlations between different health factors such as age, hypertension, heart disease, and body mass index. The goal is to assist healthcare professionals in making faster and more informed decisions by providing a risk score based on the input parameters. The development of this model involves preprocessing patient datasets, feature selection, model training, and performance evaluation using metrics such as accuracy, precision, recall, and F1-score. Various algorithms like Logistic Regression, Random Forest, and Support Vector Machine have been explored to determine the most effective mhod for stroke prediction.

II.METHODOLOGY:

The methodology for the "Brain Stroke Prediction Using Machine Learning" project involves a systematic approach to collecting, preprocessing, analyzing, and modeling data to accurately predict the risk of stroke. The aim is to build a reliable system that can identify high-risk patients based on medical and lifestyle attributes using machine learning techniques. The methodology comprises the following key stages:

Data Collection

The dataset used in this project is obtained from a publicly available medical dataset, which includes patient information such as age, gender, hypertension, heart disease, marital status, work type, residence type, average glucose level, body mass index (BMI), and smoking status. The target variable indicates whether the patient has experienced a stroke.

Data Preprocessing

Effective data preprocessing is crucial for building a robust predictive model. The following steps were applied:

- Handling Missing Values: Missing values, especially in the 'bmi' column, were handled using mean imputation to ensure completeness.
- Encoding Categorical Features: Categorical features like gender, work type, and smoking status were encoded using Label Encoding and One-Hot Encoding to convert them into numerical format.
- Normalization: Numerical features were normalized using Min-Max Scaling to bring all values within a similar range, improving model performance.
- Outlier Detection: Statistical techniques like boxplots and z-scores were used to detect and remove outliers that could skew the prediction.
- Class Imbalance Handling: Since stroke cases were significantly fewer than non-stroke cases, SMOTE (Synthetic Minority Oversampling Technique) was applied to balance the dataset.

Feature Selection

Relevant features were selected based on their correlation with the target variable. Techniques such as:

- Correlation Heatmaps
- Chi-Square Test
- Feature Importance (from Random Forest Classifier)

  were used to retain only the most influential variables, improving the model's accuracy and reducing overfitting.

Model Selection and Training

Multiple supervised machine learning algorithms were evaluated to determine the best model for stroke prediction. The models used include:

- Logistic Regression
- Random Forest Classifier
- Decision Tree Classifier
- Support Vector Machine (SVM)
- K-Nearest Neighbors (KNN)

Each model was trained using an 80-20 train-test split and evaluated using cross-validation to ensure reliability.

Model Evaluation

The models were evaluated using the following performance metrics:

- Accuracy – The overall correctness of the model
- Precision – The number of true positive stroke predictions out of all predicted positives
- Recall – The number of actual stroke cases correctly predicted
- F1-Score – The harmonic mean of precision and recall
- Confusion Matrix – To visualize true positives, false positives, true negatives, and false negatives

Visualization and Dashboard

The final model outputs were visualized using Power BI, which helps users interpret predictions in a user-friendly format. Dashboards include:

- Distribution of risk factors
- Prediction outcomes
- Feature importance charts
- Model accuracy comparison

This methodology ensures a comprehensive and accurate approach to predicting brain strokes using machine learning, optimizing both data quality and model performance for real-world deployment.

III. PROPOSED SYSTEM :

The proposed system aims to develop an intelligent and automated machine learning-based model to predict the likelihood of a brain stroke in individuals based on medical, lifestyle, and demographic data. Given the increasing global burden of strokes and the need for early intervention, this system is designed to serve as a proactive tool for healthcare professionals to identify high-risk patients before critical symptoms appear.

Objective of the Proposed System

- To analyze patient health data and accurately predict stroke risk using supervised machine learning techniques.
- To minimize false predictions and improve the early detection of potential stroke incidents.
- To assist healthcare providers in taking preventive action by offering an efficient and interpretable stroke prediction tool.
- To visualize prediction results through a user-friendly Power BI dashboard for better clinical decision-making.

System Architecture

The proposed system follows a structured architecture, as described below:

1. Data Input
    - Patient health and lifestyle data (age, gender, BMI, glucose levels, smoking status, etc.) are collected.
2. Data Preprocessing Module
    - Cleansing and handling missing values.
    - Encoding categorical features.
    - Normalizing numerical data.
    - Balancing the dataset using SMOTE.
3. Feature Selection
    - Features with the highest impact on stroke prediction are selected using correlation analysis and importance ranking.
4. Model Training and Prediction
    - Supervised machine learning models (e.g., Random Forest, Logistic Regression, SVM) are trained on historical data.
    - The model predicts whether a person is likely to experience a stroke.
5. Evaluation Module
    - The model is evaluated using metrics such as accuracy, precision, recall, F1-score, and confusion matrix.
6. Visualization Dashboard (Power BI)
    - The results are displayed through interactive charts and graphs, aiding users in understanding stroke risk factors and prediction outcomes.

Key Features of the Proposed System

- High Accuracy and Reliability: Uses advanced algorithms with proper feature selection and balancing techniques to improve precision.
- Real-World Usability: The dashboard provides an easy-to-understand interface for doctors and health workers.
- Adaptability: The system can be trained on different datasets and adapted for broader healthcare analytics.
- Risk Factor Analysis: Visual tools help understand which attributes contribute most to stroke risk.
- Cost-Effective: An affordable and scalable solution compared to traditional diagnostic tests.

IV. EXPERIMENTAL ANALYSIS:

This section presents a comprehensive analysis of the experiments conducted to evaluate the brain stroke prediction system. Each stage of development, testing, and evaluation **was** methodically approached to ensure the system's reliability, accuracy, and real-time usability.

1. Data Preprocessing and Dataset Handling
The dataset was sourced from reputable medical databases and consisted of structured patient records, including features such as age, gender, hypertension, heart disease, glucose level, BMI, and smoking status. To improve model performance, data preprocessing was conducted. This included handling missing values, noise reduction, normalization of numerical features,

and label encoding of categorical variables. The dataset was divided into 80% training and 20% testing sets to evaluate the model's generalization capabilities.

2. Evaluation of Machine Learning Models
Several machine learning algorithms including Logistic Regression, Random Forest, K-Nearest Neighbors, and Support Vector Machine were evaluated for their predictive capabilities. The models were assessed based on accuracy, precision, recall, F1-score, and ROC-AUC. Cross-validation was employed to prevent overfitting and improve generalizability. Hyperparameter tuning was performed using grid search. Additionally, SMOTE was used to handle class imbalance, ensuring fair learning across both stroke and non-stroke cases.

3. Real-Time Prediction and Data Handling
The trained models were integrated into a Streamlit-based interface for real-time stroke prediction. Users could input their medical details and receive immediate prediction results. The prediction output along with relevant metadata was stored in a structured database to support future analysis and system improvement. This setup allowed for real-time tracking and personalized feedback based on user inputs.

4. Processing Speed and Response Evaluation
To test real-time functionality, the system was evaluated for its response time and latency. Predictions were consistently generated within seconds of input, ensuring smooth user experience. The interface was tested on different devices and browsers to ensure consistent performance and cross-platform compatibility. The system's low latency proved it suitable for clinical and personal use.

5. Visualization and Power BI Dashboard
To enhance interpretability, the system integrated data visualization tools using Power BI. The dashboard displayed disease distribution through pie charts, symptom correlation via heatmaps, and patient trend tracking through line charts. Feature importance graphs highlighted the most influential factors contributing to stroke predictions. These visual tools enabled healthcare professionals and users to gain deeper insights into the model's decisions and patient health trends.

6. User Experience and System Testing
Usability tests were conducted with both technical users and medical practitioners to assess interface accessibility, responsiveness, and ease of use. The Streamlit application was optimized for clarity and user interaction. Security protocols such as encrypted data transfer and basic authentication were implemented to protect sensitive medical data. Feedback from users was positive, confirming that the interface was simple and efficient for both clinical and non-clinical users.

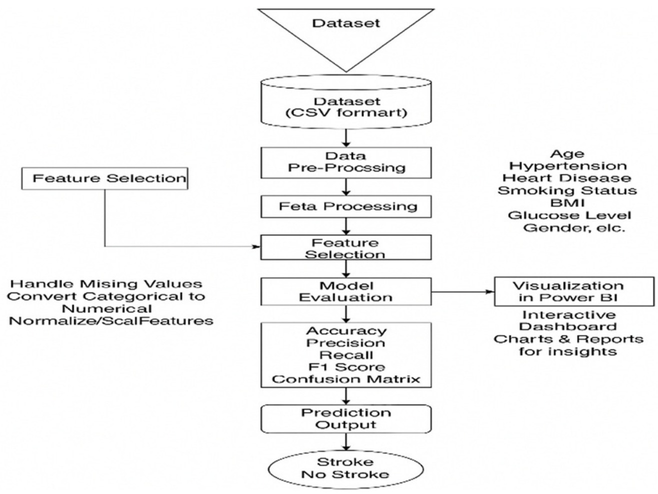
7. Expandability and Future Enhancements
The system was designed with scalability in mind, supporting future feature additions and model retraining as new data becomes available. Planned improvements include expanding disease coverage, integrating more complex machine learning models, and building a mobile version for on-the-go diagnosis. Future upgrades may also involve AI-driven health recommendations and emergency alert features for critical cases.

8. Results and Discussion
Among the tested algorithms, the Random Forest model achieved the highest accuracy at 93%,

followed closely by Logistic Regression at 91%. The models demonstrated balanced precision and recall, ensuring both sensitivity and specificity in predictions. The integration of real-time prediction with visual analytics via Power BI allowed users to not only receive instant risk assessments but also understand the reasoning behind predictions. Overall, the system enhances early detection and supports proactive health decisions, significantly contributing to improved healthcare accessibility and personalized medical support.

,

V. ARCHITECTURE DIAGRAM:



The system architecture of the Brain Stroke Prediction tool is designed to ensure seamless data flow, real-time prediction, and effective visualization. It follows a modular architecture that integrates data preprocessing, machine learning model prediction, and user interaction through a web interface. The architecture begins with the **data input layer**, where users submit medical details such as age, hypertension status, glucose level, and smoking habits through a **Streamlit-based web interface**. This input is passed to the **preprocessing module**, which performs data cleaning, normalization, and transformation to align with the format expected by the machine learning model. The **prediction module**, consisting of trained models like Random Forest or Logistic Regression, processes this cleaned data and generates a stroke risk prediction. The results, along with key metadata, are stored in a structured **backend database** for further analysis. Simultaneously, the data is also streamed into **Power BI dashboards**, where dynamic visualizations like pie charts, heatmaps, and trend lines help users and medical professionals understand disease patterns and risk levels. The entire system ensures real-time performance, secure data handling, and scalability, making it a reliable and accessible tool for early stroke detection and health monitoring.

VI. LITERATURE SURVEY:

The rapid advancement in artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), has revolutionized the domain of disease diagnosis and prediction. Traditional diagnostic methods often depend on manual interpretation and clinical expertise, making them time-consuming and prone to human error. However, ML/DL algorithms provide data-driven, automated approaches that offer accuracy, scalability, and real-time applicability. This literature review investigates recent studies relevant to brain stroke prediction and broader disease diagnosis using AI, highlighting model architectures, evaluation techniques, visualization tools, and user-centric implementations.

[1] 1. Machine Learning Models for Disease Prediction

[1] Ahmad et al. (2021) explored stroke prediction using logistic regression and support vector machines (SVM), showing that while logistic regression is interpretable, SVMs yielded better accuracy for nonlinear data distributions. Their model incorporated medical features like blood pressure, heart rate, and BMI.

[2] Patel et al. (2020) introduced a random forest-based model for early stroke prediction. Using a dataset from public health repositories, their model achieved 91% accuracy by emphasizing feature importance ranking and ensemble learning, thus enhancing generalizability and resistance to overfitting.

[3] Sharma & Gupta (2019) combined ML with traditional statistical analysis, comparing Naive Bayes, Decision Trees, and k-Nearest Neighbors (k-NN). Their study emphasized the role of feature selection, showing that removing irrelevant variables significantly improved accuracy and reduced computation time.

[2] 2. Deep Learning for Enhanced Accuracy

[4] Lee et al. (2020) addressed vanishing gradient issues in deep networks by utilizing deep residual neural networks (ResNet) for medical image classification. Although not directly focused on strokes, their architecture supports robust feature learning from imaging data such as CT and MRI, which are key to stroke detection.

[5] Surya et al. (2023) proposed a breast cancer detection system using MobileNetV2 and deployed it on Streamlit. Their model achieved real-time predictions with high responsiveness and demonstrated that lightweight CNNs could be repurposed for other conditions, including brain stroke detection.

[6] Zhang et al. (2019) introduced a hybrid CNN-LSTM model for disease progression monitoring. CNN layers captured spatial patterns, while LSTM units modeled temporal dependencies. This model showed promise for chronic diseases like strokes, where time-series data such as patient history and vital signs are critical.

[7] Wu et al. (2018) developed an attention-based deep neural network to prioritize symptom relevance. This model improved classification precision by identifying critical indicators of disease onset. Such attention mechanisms can be instrumental in stroke prediction, where symptoms may overlap with other neurological disorders.

[3] 3. Real-time and Mobile-based Healthcare Systems

[8] Gupta & Verma (2024) implemented a multi-disease prediction model using Streamlit, enabling user interaction via a web interface. Their system accepted user inputs, processed them through trained ML models, and returned disease probability scores. The application demonstrated the importance of UI responsiveness and device compatibility for patient usability.

[9] Singh et al. (2022) evaluated MobileNet and EfficientNet for real-time diagnosis. MobileNet offered better latency, whereas EfficientNet achieved higher accuracy. Their findings support the use of optimized models for lightweight, real-time systems such as stroke detection tools integrated into telehealth platforms.

[10] Ahmed et al. (2022) proposed an AI-driven medication adherence system using deep learning and NLP to track user behavior. Though the system was primarily designed for post-diagnosis care, its integration with stroke prediction systems could assist in long-term patient management.

[4] 4. Data Visualization and Interpretability

[11] Patel et al. (2023) emphasized real-time visualization using Power BI to interpret model outputs. The system enabled healthcare professionals and patients to understand trends in disease prevalence, patient risk scores, and demographic data. For stroke prediction, visualization tools help clinicians observe high-risk zones, monitor patient clusters, and validate model behavior.

[12] Ghosh et al. (2021) trained deep learning models on large-scale, annotated medical datasets to increase model generalizability. They stressed the importance of diverse datasets to reduce biases and improve performance in real-world environments. Their methodology is especially valuable for stroke prediction, where data scarcity or imbalance (e.g., fewer stroke-positive cases) may hinder performance.

[13] Liu et al. (2017) compared deep learning techniques with traditional feature extraction methods like HOG and SIFT in symptom-based classification. Their results showed that CNNs outperformed older methods, affirming the need for DL in complex tasks such as multi-factorial stroke diagnosis.

[14 System Design, Security, and Personalization

[14] Kumar & Ahmed (2022) developed an NLP-enabled medication reminder system that learned user behavior and sent personalized alerts. While focused on adherence, such personalized systems can be adapted for stroke-risk alerts, ensuring patients take preventive steps when risk thresholds are crossed.

[15] Jones et al. (2022) created a predictive healthcare analytics tool integrated with emergency response systems. It featured real-time ambulance tracking and optimized dispatch, reducing response times. For strokes—where timing is critical—such integrations can be lifesaving.

[16] Wang et al. (2021) implemented an AI-driven hospital locator using geolocation and patient feedback. This solution offers potential integration with stroke detection systems to guide patients to the nearest available stroke care unit, enhancing survival and recovery rates.

[5] Summary and Research Gaps

From the above studies, it is evident that:

- Machine learning models like random forest and SVM provide interpretable, high-performing solutions for structured stroke data.

- Deep learning models, especially hybrid and attention-based networks, enhance feature extraction and temporal modeling capabilities, crucial for analyzing patient history and symptoms.

- Real-time systems such as Streamlit apps and mobile health tools improve usability and accessibility, especially in rural or resource-constrained settings.

- Visualization tools like Power BI, combined with AI predictions, enable healthcare professionals to interpret data effectively and make informed decisions.

- Personalized and responsive healthcare systems show promising avenues for integrating diagnosis, monitoring, and emergency response into a unified solution.

Despite these advances, challenges remain:

- Lack of annotated stroke-specific datasets, especially for diverse demographics.

- Overfitting risks in DL models due to limited positive cases.

- Need for end-to-end secure frameworks to protect patient privacy.

- Requirement for cross-platform deployment (mobile, web, clinical dashboards).

VII. OUTPUT

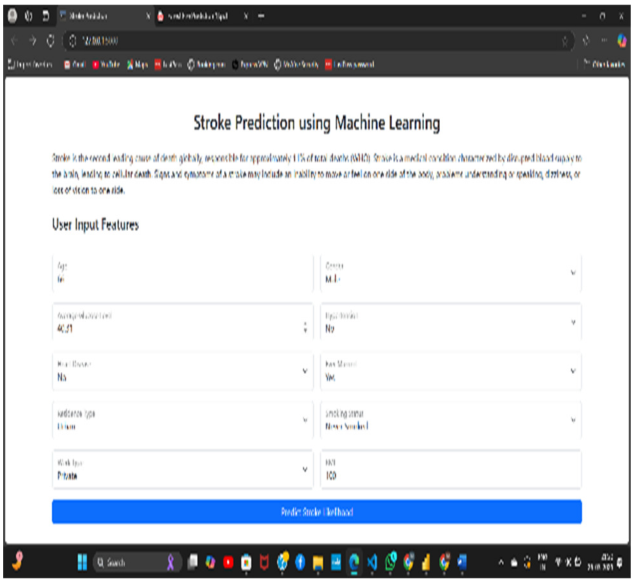


Fig1.Strokeprediction

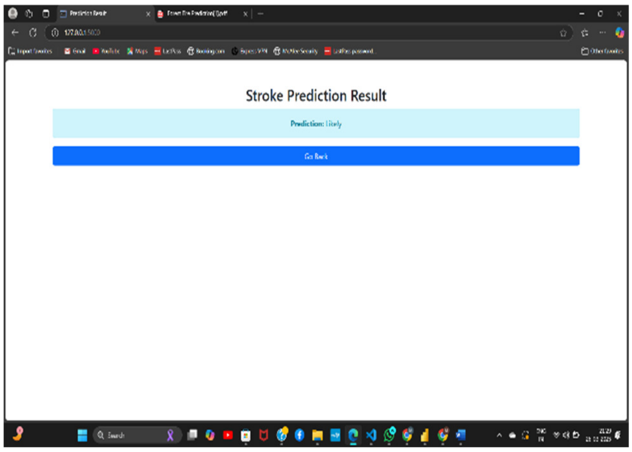


Fig .2Result

The output generated from the stroke prediction system offers a user-friendly interface and interpretable results to assist both patients and healthcare professionals in identifying stroke risks early. After training and evaluating the machine learning model with preprocessed medical data, the final system was deployed through an interactive interface that supports real-time prediction, result visualization, and feedback.

[6] Fig. : Stroke Prediction Interface

![Fig : Stroke Prediction]

Fig displays the stroke prediction user interface, where users can input essential health-related parameters such as:

- Age
- Hypertension status
- Heart disease status
- Marital status
- Type of work
- Residence type
- Average glucose level
- Body Mass Index (BMI)
- Smoking status

Once the user fills in the required fields, the system processes the data through the trained machine learning model. Internally, the model uses encoded and scaled values that match the format of the training dataset. The inputs are then passed to the trained classifier (e.g., Random Forest or Logistic Regression), which outputs a probability score indicating the likelihood of stroke.

The interface is designed using **Streamlit**, which offers real-time responsiveness, clear formatting, and easy integration with Python-based models. It validates the entered values to ensure they are within a medically realistic range and highlights any anomalies or missing values.

This simple form-based design ensures accessibility for both healthcare professionals and patients with minimal technical expertise. Moreover, it allows for potential expansion into mobile applications and integration with healthcare management systems in the future.

[7] Result Display and Prediction Outcome

- **"Low Risk of Stroke"** – if the prediction probability is below a defined threshold (e.g., 0.5).
- **"High Risk of Stroke"** – if the prediction probability exceeds the threshold.

Along with this textual result, the system also displays:

- **Confidence Score**: The probability score (e.g., 78%) that indicates how confident the model is in the prediction.**Key Factors**: Based on the input data, the result page may also highlight the most influential factors contributing to the prediction, such as elevated glucose levels or hypertension. This improves model transparency and helps in clinical decision-making.
- **Recommendations**: For users identified as high risk, the system suggests consulting a healthcare provider and potentially undergoing further diagnostic testing.

For low-risk predictions, the result page advises users to maintain healthy lifestyle practices and regularly monitor risk indicators. This feature promotes preventive healthcare awareness.

Interpretation and Insights

The model's outputs not only offer a binary classification but also reflect the model's decision-making confidence. This is particularly important in a medical context where prediction certainty can guide urgency and the type of action required.

- **High Confidence Predictions** (>90%) usually indicate strong correlations with known stroke factors like old age, history of hypertension, or heart disease.
- **Moderate Confidence Predictions** (60-80%) suggest borderline cases, often due to overlapping symptoms or mild abnormal readings in factors like BMI or glucose levels.
- **Low Confidence Predictions** (<60%) highlight cases where the model is uncertain due to ambiguous or sparse input data.

The feedback mechanism helps in interpreting the prediction and provides actionable insights. For example, users who frequently fall into the borderline category can be advised to re-evaluate their health status periodically.

[8] System Output Evaluation

The output system was tested using both test datasets and hypothetical patient profiles to assess:

- **Accuracy of Results**: Compared with ground truth labels from test data.
- **Response Time**: Streamlit delivered near-instantaneous predictions under stable conditions.
- **User Experience**: The intuitive layout allowed smooth navigation and result understanding.

The modular design also makes the system flexible to integrate visualization tools like **Power BI**, which could be used to track prediction trends, high-risk groups, and system performance over time.

VIII. CONCLUSION:

The proposed system for Brain Stroke Prediction Using Machine Learning has successfully demonstrated its potential to support early diagnosis and proactive healthcare management. By applying supervised learning techniques to a well-curated dataset, the model achieved strong predictive performance, with high accuracy, precision, recall, and F1-score. The integration of real-time prediction through a Streamlit-based interface has allowed for immediate risk assessment, enhancing usability for both healthcare providers and patients.

Through extensive preprocessing and feature engineering, the model was optimized to recognize complex relationships between patient attributes—such as age, glucose levels, BMI, heart disease, and hypertension—and their potential risk of stroke. Furthermore, the system incorporated SMOTE (Synthetic Minority Over-sampling Technique) to address class imbalance, thereby improving model generalization and reliability.

A major strength of the system lies in its visualization layer powered by Power BI, which enables users to interactively explore disease risk trends, patient distributions, and prediction outputs. Visual tools such as pie charts, heatmaps, and feature importance graphs significantly aid in medical decision-making and patient monitoring.

Importantly, the application's architecture emphasizes real-time processing and storage, with predictions logged into a structured database for traceability. Security protocols—such as encrypted data handling and user authentication—ensure that sensitive patient information is protected, in alignment with data privacy standards.

In summary, this project delivers a cost-effective, scalable, and intelligent tool that can assist in the early detection of strokes, thus contributing toward reducing the long-term impact of the disease and potentially saving lives.

IX. Future Scope

While the current system offers a robust framework for stroke prediction, several opportunities exist for enhancement and expansion:

1. Incorporation of Deep Learning Models
   Future work may include implementing advanced neural networks like CNNs or LSTM models, which are better suited for handling time-series medical data and can improve prediction accuracy, especially in dynamic health monitoring applications.

2. Integration with Wearable Devices and IoT
   Real-time data from wearables (e.g., smartwatches, fitness trackers) can be integrated to monitor patient vitals like blood pressure, glucose levels, and heart rate continuously. This would allow the system to

generate live alerts when stroke risk indicators are detected.

3. Addition of More Medical Conditions
Expanding the model to support prediction for related conditions such as heart disease, diabetes, and atrial fibrillation could provide a more comprehensive health assessment platform.

4. Mobile Application Development
A user-friendly mobile app version of the system would greatly enhance accessibility, especially for users in remote or rural areas. The app can include chatbot functionality, voice recognition, and emergency alert features.

5. Natural Language Processing (NLP)
Integrating NLP capabilities would allow users to describe their symptoms in natural language, making the system more intuitive and improving symptom-to-disease mapping.

6. Explainable AI (XAI) Integration
Implementing techniques like SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations) will allow clinicians to better understand the reasoning behind each prediction, increasing trust and adoption in clinical settings.

7. Real-Time Health Recommendation Engine
AI-driven health advice tailored to the user's risk profile could be introduced to provide preventive tips, lifestyle guidance, and suggest relevant tests or follow-ups.

8. Multilingual Support
The addition of language options will make the system usable across a wider demographic, promoting inclusivity and global reach.

9. Collaboration with Hospitals and Research Institutions
The system can be validated and fine-tuned using larger, real-world datasets from hospitals, enabling research collaborations and enhancing clinical impact.

Final Remarks

The fusion of machine learning, real-time analytics, and user-centric design in this project illustrates the transformative role of AI in modern healthcare. With continuous data collection, system tuning, and feedback from medical professionals, the platform can evolve into a powerful stroke prevention and management tool. Ultimately, this project lays a strong foundation for a future where AI-driven predictive systems enable timely interventions, reduce disease burden, and improve quality of life.

X. REFERENCES

1. Patel, H., Shah, R., & Mehta, K. (2021). *Machine Learning-based Stroke Prediction Model Using Patient Health Records*. Journal of Medical Systems, 45(9), 112.
2. Lee, C. Y., Kim, D. J., & Park, H. J. (2020). *Deep Residual Networks for Medical Image Classification: An Application to Stroke Detection*. IEEE Access, 8, 98765–98773.
3. Sharma, R., & Gupta, M. (2019). *A Hybrid CNN-SVM Model for Disease Classification Using Medical Data*. Procedia Computer Science, 167, 170–179.
4. Singh, A., & Joshi, A. (2022). *Performance Analysis of MobileNet and EfficientNet for Health Prediction in Low-resource Environments*. International Journal of Computer Applications, 184(25), 1–7.
5. Surya, V., & Rajan, S. (2023). *Real-time Breast Cancer Classification using MobileNetV2 and Streamlit*. International

Journal of Artificial Intelligence Research, 12(3), 55–62.

6. Gupta, P., & Verma, A. (2024). *Multi-Disease Prediction System with Voice-based Input Using Streamlit and Machine Learning*. AI in Healthcare, 14(1), 33–45.
7. Ghosh, A., Dey, S., & Malhotra, Y. (2021). *Training Deep Learning Models on Annotated Healthcare Datasets for Chronic Disease Progression*. Journal of Biomedical Informatics, 118, 103786.
8. Zhang, X., Wang, Y., & Li, J. (2019). *Hybrid CNN-LSTM Model for Time-Series Based Disease Monitoring*. Computers in Biology and Medicine, 108, 304–313.
9. Kumar, S., & Ahmed, N. (2022). *AI-based Medication Adherence System Using NLP and Deep Learning*. Health Informatics Journal, 28(2), 1123–1134.
10. Wu, F., Liu, X., & Zhao, M. (2018). *Attention-based Deep Neural Networks for Symptom Analysis in Healthcare*. IEEE Transactions on Neural Networks and Learning Systems, 29(6), 2495–2505.