Anusandhanvallari
Vol 2026, No.1
July 2026
ISSN 2229-3388

Brain Stroke Detection and Doctor Recommendation System Using Machine Learning: An Integrated Approach to Predictive Healthcare

[1]Amrita Dubey, [2]Kunal Goenka, [3]Anshita Soni, [4]Dr. Chaitali Biswas Dutta

[1]Department of Computer Science and Engineering Shri Shankaracharya Institute of Professional Management & Technology, Raipur, Chhattisgarh, India

[2]Department of Computer Science and Engineering Shri Shankaracharya Institute of Professional Management & Technology, Raipur, Chhattisgarh, India

[3]Department of Computer Science and Engineering Shri Shankaracharya Institute of Professional Management & Technology, Raipur, Chhattisgarh, India

[4]Associate Professor Department of Computer Science and Engineering Shri Shankaracharya Institute of Professional Management & Technology, Raipur, Chhattisgarh, India

Abstract—Brain stroke remains one of the foremost causes of death and permanent neurological disability worldwide, imposing substantial socioeconomic burdens on healthcare systems globally. Timely and accurate prediction of stroke risk, combined with seamless referral to qualified specialists, is clinically decisive in minimizing irreversible brain damage and improving patient outcomes. This paper presents a comprehensive integrated Brain Stroke Detection and Doctor Recommendation System that unifies a machine learning–based diagnostic engine with a web-based intelligent specialist referral platform. The stroke detection module employs a Random Forest (RF) Classifier trained on the publicly available Kaggle Stroke Prediction Dataset comprising 5,110 patient records across 11 clinical and demographic features. A rigorous comparative study is conducted against Logistic Regression and Decision Tree classifiers under identical experimental conditions. The RF model achieves an overall accuracy of 95.66%, demonstrating robust performance characterized by high precision on the majority (non-stroke) class, a non-trivial challenge acknowledged by the severe class imbalance (4.9% stroke prevalence) inherent to real-world medical datasets. Upon detection of high stroke risk, the system automatically activates a doctor recommendation module that ranks neurologists and stroke specialists using a three-component hybrid algorithm combining content-based filtering, collaborative filtering, and ontology-based medical reasoning, further augmented by geolocation-based proximity scoring. The full system is implemented using Python Flask, PHP/MySQL, and a responsive HTML/CSS/JavaScript frontend, deployed locally on XAMPP for offline accessibility. Experimental results confirm that the proposed end-to-end architecture effectively bridges the critical translational gap between automated AI-based diagnosis and actionable patient care navigation, with potential for deployment in both urban hospital settings and resource-constrained rural healthcare environments.

Index terms—Brain stroke prediction, Random Forest Classifier, doctor recommendation system, machine learning, healthcare decision support, hybrid filtering, content-based filtering, collaborative filtering, clinical data analysis, class imbalance, Flask, telemedicine.

I. Introduction

Stroke is a life-threatening neurological emergency that occurs when the blood supply to a region of the brain is abruptly interrupted, either due to arterial blockage (ischemic stroke, comprising approximately 87% of all cases) or rupture of a blood vessel causing hemorrhage (hemorrhagic stroke). The World Health Organization (WHO) ranks stroke as the second leading cause of death globally and the third leading cause of disability-adjusted life years (DALYs) lost [1]. According to the Global Burden of Disease (GBD) study, approximately 12.2 million new stroke cases occurred worldwide in 2019, with 6.55 million deaths attributable to stroke [2]. In India

specifically, stroke has emerged as a major public health crisis, with an estimated incidence rate of 105–152 per 100,000 person-years in rural populations and 334–424 per 100,000 person-years in urban populations [3].

The neurological damage caused by stroke is profoundly time-dependent. Saver [4] demonstrated that for every minute of delayed reperfusion in ischemic stroke, approximately 1.9 million neurons, 13.8 billion synapses, and 12 km of axonal fibers are irreversibly lost. This biological imperative underlies the clinical maxim "time is brain" and underscores the paramount importance of early detection and rapid specialist intervention. Despite advances in acute stroke care including thrombolytic therapy and mechanical thrombectomy, delayed diagnosis and insufficient access to specialized neurological care continue to be major barriers, particularly in semi-urban and rural regions of low- and middle-income countries (LMICs) where neurologist density may be as low as 0.03 per 100,000 population [5].

Recent years have witnessed an exponential growth in the application of Artificial Intelligence (AI) and Machine Learning (ML) techniques in healthcare. Supervised classification algorithms trained on structured electronic health records (EHR) and clinical datasets have demonstrated remarkable capability in predicting disease onset and progression across multiple domains including cardiovascular disease, diabetes, cancer, and neurological disorders. Compared to conventional diagnostic protocols that rely on symptom recognition and clinician judgment, ML-based risk stratification tools offer the potential for population-level, real-time screening using routinely collected health parameters, enabling proactive identification of at-risk individuals before clinical manifestation.

However, a critical and largely unaddressed gap persists in the existing literature and deployed systems: the failure to integrate the prediction output with an immediate, actionable, patient-centered response. Most existing stroke prediction tools are standalone classification models that output a risk score or label but provide no guidance on what the patient should do next. In the absence of automated referral, patients must independently navigate the complex healthcare ecosystem to locate and contact appropriate specialists—a process that introduces dangerous delays, especially for individuals with limited health literacy or geographic access constraints.

This paper addresses this gap by proposing and implementing an integrated two-module architecture that transforms the paradigm from passive risk reporting to active healthcare navigation. The system comprises: (1) a Stroke Detection Module employing a comparative evaluation of Logistic Regression, Decision Tree, and Random Forest classifiers with the RF model selected for deployment based on empirical performance analysis; and (2) a Web-Based Doctor Recommendation Module utilizing a three-component hybrid filtering algorithm (content-based, collaborative, and ontology-driven) augmented by geolocation-based proximity scoring for real-time specialist referral.

The principal contributions of this work are summarized as follows:

1) A comprehensive and fair comparative evaluation of three supervised ML algorithms—Logistic Regression, Decision Tree, and Random Forest—on the Kaggle Stroke Prediction Dataset, including rigorous analysis of class-imbalance effects on minority-class detection performance and model selection rationale.
2) A novel hybrid recommendation engine that combines content-based filtering for clinical feature similarity matching, collaborative filtering based on historical patient feedback, and ontology-based medical reasoning for context-aware specialist-type routing.
3) A detailed feature engineering and preprocessing pipeline tailored to the challenges of clinical stroke data, including missing value imputation, categorical encoding, Min-Max normalization, and class imbalance mitigation strategies.

4) A fully functional, end-to-end integrated system implementation using Python Flask, PHP/MySQL, and a responsive web interface, with local deployment on XAMPP for accessibility in connectivity-limited environments.

5) A thorough discussion of system limitations, ethical considerations, clinical translation challenges, and a concrete roadmap for future enhancements including IoT integration, deep learning extensions, and cloud deployment.

The remainder of this paper is organized as follows. Section II provides a comprehensive review of the related literature on ML-based stroke prediction and healthcare recommendation systems. Section III describes the dataset characteristics and the complete preprocessing pipeline. Section IV presents the theoretical foundations and architectural details of both modules. Section V reports experimental results with detailed statistical analysis. Section VI provides an in-depth discussion of findings, limitations, and ethical considerations. Section VII concludes the paper with future directions.

II. Related Work

A. Machine Learning for Stroke Risk Prediction

The application of ML to stroke risk prediction has been an active area of research since the early 2010s, with a notable acceleration in publication volume following the widespread availability of the Kaggle Stroke Prediction Dataset. Early work in this domain primarily relied on traditional statistical models. Logistic Regression was a common choice due to its interpretability and computational efficiency, but its assumption of linear separability between classes proved limiting in the presence of complex, nonlinear interactions among clinical risk factors such as glucose levels, BMI, and age [6].

The transition toward ensemble learning methods marked a significant improvement in predictive performance. Breiman [7] introduced the Random Forest algorithm as an ensemble of independently trained decision trees, demonstrating superior accuracy and variance reduction compared to single-tree classifiers. Subsequent healthcare applications rapidly adopted RF due to its robustness on heterogeneous medical data, its ability to handle missing values, and its built-in feature importance estimation—a clinically valuable property for understanding which health factors most strongly contribute to stroke risk.

Sharma et al. [8] conducted a comparative study of Random Forest, Decision Tree, and XGBoost on a stroke prediction dataset derived from real clinical records, demonstrating that ensemble models consistently outperformed single-tree classifiers in terms of accuracy and recall. Their study highlighted the predictive importance of BMI, average glucose level, smoking status, and age, confirming the clinical literature. Liu et al. [9] further demonstrated that the quality of data preprocessing—particularly missing value imputation strategies and feature normalization techniques—has a comparable or greater impact on final model performance than the choice of algorithm itself, underscoring the importance of rigorous pipeline design.

Alkhateeb et al. [10] extended the feature space to include behavioral indicators and proposed hybrid deep learning models based on Multi-Layer Perceptrons (MLPs) for multi-dimensional stroke risk assessment. Their work demonstrated the capacity of neural architectures to capture deeply nonlinear interdependencies among clinical features that shallower models may miss. Khan et al. [11] proposed ensemble voting classifiers that aggregate predictions from RF, SVM, and k-NN models, achieving improved robustness and reduced prediction variance. Critically, they also raised the issue of model interpretability as a prerequisite for clinical adoption, advocating for Explainable AI (XAI) approaches in medical decision support systems.

A fundamental challenge across all of these studies is the handling of class imbalance. Stroke is a relatively rare event in population-level datasets, resulting in training data where non-stroke cases vastly outnumber stroke cases. This imbalance causes classifiers to develop a systematic bias toward the majority class, resulting in artificially

high overall accuracy but critically low recall on the clinically important minority (stroke) class. Techniques to address this include oversampling (SMOTE—Synthetic Minority Over-sampling Technique), undersampling, and class-weighted loss functions [12].

B. Healthcare Recommendation Systems

Healthcare recommendation systems have been studied in the broader context of decision support for patient-provider matching. Content-based filtering approaches match patients to healthcare providers based on clinical needs and provider specialization profiles. Collaborative filtering methods leverage historical interaction data—such as patient ratings and treatment outcomes—to recommend providers with proven effectiveness for similar patient profiles [13].

Kumar and Singh [14] surveyed doctor recommendation models across multiple dimensions including geographic proximity, specialization matching, and patient feedback integration. Their work identified a critical limitation common to most systems: they operate as isolated recommendation platforms, disconnected from any diagnostic or risk prediction tool. As a result, patients receive no automated guidance following risk identification. Li and Chen [15] proposed an integrated framework coupling a cardiovascular risk model with a hospital recommendation system, demonstrating the feasibility and clinical value of tight prediction-referral integration, though their system was not validated on stroke-specific populations.

Mobile health (mHealth) applications have emerged as a consumer-facing platform for bridging the gap between risk awareness and care access. However, Patel et al. [16] observed that most mHealth applications lack AI-driven predictive analytics, relying instead on symptom checklists or static risk calculators. Furthermore, commercial platforms typically depend on third-party mapping APIs that are not locally deployable, rendering them unusable in environments with unreliable internet connectivity.

C. Research Gap and Positioning

A synthesis of the literature reveals a consistent and significant gap: while considerable progress has been made in both stroke prediction accuracy and healthcare recommendation methodology, the two streams have developed largely in isolation. Existing systems either predict stroke risk without recommending a course of action, or recommend healthcare providers without incorporating any diagnostic intelligence. The present work is positioned to bridge this gap by developing a tightly integrated architecture in which the prediction and recommendation modules operate as a unified clinical decision support pipeline.

III. Dataset Description And Preprocessing

A. Dataset Overview

The dataset used in this study is the publicly available Kaggle Stroke Prediction Dataset [17], which has been extensively used in the stroke prediction literature as a standardized benchmark. The dataset contains 5,110 patient records, each described by 11 features encompassing clinical, demographic, and lifestyle attributes. The target variable is a binary stroke label (1 = stroke, 0 = no stroke). Table I provides a detailed description of all features.

Table I. Description Of Dataset Features

Feature	Type	Range / Values	Clinical Relevance
Age	Continuous	0.08 – 82	Primary risk factor; stroke incidence rises sharply above 65 years
Gender	Categorical	Male/Female/Other	Sex-specific hormonal and vascular risk profiles
Hypertension	Binary	0 / 1	Major modifiable risk factor; raises stroke risk 3–4×
Heart Disease	Binary	0 / 1	Atrial fibrillation and cardiac embolism are leading stroke causes
Ever Married	Binary	Yes / No	Proxy for age and socioeconomic status
Work Type	Categorical	5 categories	Occupational stress and activity level correlate with vascular risk
Residence Type	Categorical	Urban / Rural	Proxy for healthcare access and environmental exposures
Avg Glucose Level	Continuous	55.12 – 271.74	Elevated glucose (diabetes) is a key independent stroke risk factor
BMI	Continuous	10.3 – 97.6	Obesity linked to hypertension, diabetes, and atherosclerosis
Smoking Status	Categorical	4 categories	Doubles stroke risk through oxidative stress and platelet aggregation
Stroke (Target)	Binary	0 / 1	Class distribution: 4,861 (No) vs. 249 (Yes) — 4.87% prevalence

Table I: Feature descriptions, types, and clinical relevance for the Kaggle Stroke Prediction Dataset.

B. Exploratory Data Analysis

Exploratory Data Analysis (EDA) was conducted to understand data distributions, feature correlations, and the extent of class imbalance. Key observations include:

- Age: The mean age of stroke patients (67.7 years) was substantially higher than non-stroke patients (41.9 years), confirming age as the dominant risk factor. The age distribution in stroke cases was heavily right-skewed toward older cohorts.
- Glucose Level: Stroke patients showed a bimodal distribution in average glucose levels with a pronounced secondary peak above 200 mg/dL, consistent with the known association between diabetes and stroke risk.

- BMI: The BMI attribute contained 201 missing values (3.93% of records), requiring imputation. The distribution was approximately normal with a mean of 28.9 kg/m².
- Hypertension & Heart Disease: 26.6% of stroke patients had hypertension compared to 9.3% of non-stroke patients. Similarly, 18.9% of stroke patients had heart disease compared to 4.8% of non-stroke patients.
- Class Imbalance: The target variable exhibited severe class imbalance, with only 249 stroke cases (4.87%) out of 5,110 total records. This imbalance is the primary challenge for minority-class prediction performance.

C. Data Preprocessing Pipeline

A systematic five-step preprocessing pipeline was designed and applied to the raw dataset prior to model training.

Step 1 – Missing Value Imputation: The 201 missing values in the BMI feature were imputed using median imputation rather than mean imputation, as the BMI distribution exhibited mild right skewness and median is more robust to outlier influence. Records with missing values in other features were found to be negligible and were removed.

Step 2 – Categorical Encoding: Nominal categorical variables (gender, work type, residence type, smoking status, marital status) were transformed into integer representations using Label Encoding. This approach is appropriate for tree-based models such as Random Forest, which do not assume ordinality in encoded categorical features.

Step 3 – Feature Normalization: Continuous numerical features (age, average glucose level, BMI) were scaled to the [0, 1] range using Min-Max normalization, defined as: $x' = (x - x_{min}) / (x_{max} - x_{min})$. This prevents features with larger absolute magnitudes from disproportionately influencing distance-based components of any downstream recommendation algorithm.

Step 4 – Class Imbalance Handling: Given the 4.87% stroke prevalence, oversampling of the minority class was explored using random oversampling to create a more balanced training distribution. This was applied exclusively to the training set after data splitting to prevent data leakage into the test set.

Step 5 – Train-Test Splitting: The preprocessed dataset was partitioned using stratified random splitting with an 80:20 ratio, yielding a training set of 4,049 records and a test set of 1,022 records (adjusted for oversampling applied to training only). Stratification ensures that the class distribution in both splits mirrors the original dataset, preventing artificially optimistic evaluation metrics.

D. Feature Importance Analysis

Following model training, the RF classifier's built-in Gini importance scores were extracted to identify the most predictive features. The results, summarized in Table II, reveal that age is by far the dominant predictor, followed by average glucose level and BMI. Binary clinical risk factors (hypertension, heart disease) contribute meaningfully but with lower individual importance, consistent with their status as necessary but not sufficient risk conditions in isolation.

Table Ii. Random Forest Feature Importance Scores

Rank	Feature	Gini Importance Score
1	Age	0.2841
2	Average Glucose Level	0.1923
3	BMI	0.1756

4	Hypertension	0.0892
5	Heart Disease	0.0741
6	Smoking Status	0.0634
7	Work Type	0.0512
8	Ever Married	0.0389
9	Gender	0.0198
10	Residence Type	0.0114

Table II: Random Forest Gini feature importance scores (higher = more predictive).

Iv. System Methodology And Architecture

The proposed system adopts a modular, two-phase pipeline architecture. Phase I encompasses the Stroke Detection Module, responsible for ingesting patient health data, running the trained ML classifier, and generating a risk label and probability score. Phase II encompasses the Doctor Recommendation Module, which is activated conditionally upon a high-risk detection and is responsible for retrieving, ranking, and presenting relevant specialist recommendations. The overall system architecture is described below.

A. Stroke Detection Module

The Random Forest (RF) Classifier was selected as the primary detection engine following empirical comparative evaluation. The RF model is an ensemble learning method that constructs a forest of T independently trained decision trees $\{h_1, h_2, \ldots, h^T\}$, where each tree h_i is trained on a bootstrap sample B_i drawn from the training data with replacement, using a randomly selected subset of m features at each node split (typically $m = \sqrt{M}$ where M is the total number of features). For a given input feature vector x representing a patient's clinical profile, the final binary class prediction H(x) is determined by majority voting across all trees:

$$H(x) = \text{argmax}_c \, \Sigma_{i=1}^{T} \; I(h_i(x) = y), \quad y \in \{0, 1\}$$

where $I(\cdot)$ denotes the indicator function, $y = 1$ represents the stroke class, and $y = 0$ represents the non-stroke class. In addition to the class label, the model also outputs a stroke probability estimate $\hat{P}(y=1|x) = (1/T) \, \Sigma \, I(h_i(x) = 1)$, which serves as the continuous risk score used by the recommendation module to calibrate referral urgency.

Hyperparameter tuning was performed using 5-fold stratified cross-validation with grid search over the following parameter space: number of estimators n_estimators ∈ {100, 200, 300, 500}; maximum tree depth max_depth ∈ {None, 10, 20, 30}; minimum samples per leaf min_samples_leaf ∈ {1, 2, 4}; and the criterion for split quality ∈ {gini, entropy}. The macro-averaged F1-score was used as the optimization objective, specifically chosen over accuracy to avoid being misled by the imbalanced class distribution.

B. Comparative Algorithm Descriptions

Logistic Regression (LR): LR models the probability of stroke as a sigmoid transformation of a linear combination of input features: $P(y=1|x) = \sigma(w^T x + b)$, where w is the weight vector, b is the bias term, and $\sigma(\cdot)$ is the logistic sigmoid function. Model parameters are estimated by maximizing the log-likelihood using gradient descent with

L2 regularization. LR serves as the interpretable linear baseline and is well-suited for identifying individual feature contributions, but is inherently limited by its linear decision boundary.

Decision Tree (DT): DT constructs a hierarchical binary tree structure by recursively partitioning the feature space using the Gini impurity or information gain criterion. At each node, the feature-threshold pair that maximally reduces impurity is selected as the split. DTs produce human-interpretable decision rules and can model nonlinear boundaries, but are highly susceptible to overfitting when grown without depth constraints, particularly on small training sets.Random Forest (RF): As described above, RF mitigates the variance of individual DTs through ensemble averaging and random feature subsampling. The decorrelation of individual trees—achieved by training each on a different bootstrap sample with a different random feature subset—is the key mechanism by which RF achieves superior generalization performance compared to both LR and single DTs.

C. Doctor Recommendation Module

The recommendation module is designed as a responsive, event-driven subsystem triggered automatically when the detection module returns a stroke probability $\hat{P}(y=1|x)$ exceeding a configurable threshold (default: 0.5). The module architecture consists of three algorithmic components and a geolocation-based ranking layer.

Content-Based Filtering (CBF): CBF computes a feature similarity score between the patient's clinical profile and each doctor's recorded specialization attributes. Key matching dimensions include stroke type classification (ischemic vs. hemorrhagic, derived from imaging flags if available), required specialist type (neurologist vs. neurosurgeon vs. rehabilitation specialist), severity score (derived from $\hat{P}$), and patient age group. The cosine similarity between the patient's need vector and each doctor's capability vector is computed as: $\text{Simc}_B(p, d) = (p \cdot d) / (||p|| \, ||d||)$.

Collaborative Filtering (CF): CF leverages historical patient feedback stored in the MySQL database. For each doctor d, a CF similarity score is computed based on treatment outcome ratings and patient satisfaction scores from past consultations with patients sharing similar clinical profiles. This component introduces experiential quality signals that CBF alone cannot capture, such as physician bedside manner, appointment accessibility, and treatment efficacy feedback.

Ontology-Based Medical Reasoning (OMR): An ontological mapping layer encodes domain medical knowledge about stroke specialist routing. Key ontological rules include: hemorrhagic stroke cases → neurosurgeon (for surgical intervention); ischemic stroke cases → vascular neurologist (for thrombolysis); post-acute cases → rehabilitation neurologist; and general high-risk cases → general neurologist. This ensures that recommendations are not only proximity- and quality-optimized but also medically appropriate.

The composite recommendation score for doctor d given patient profile p is computed as:

$$\text{Score}(p, d) = \alpha \cdot \text{Simc}_B(p,d) + \beta \cdot \text{Simc}^f(p,d) + \gamma \cdot \text{OMR}(p,d) + \delta \cdot \text{ProxScore}(p,d)$$

where α, β, γ, and δ are non-negative weighting coefficients summing to 1.0 ($\alpha = 0.30$, $\beta = 0.25$, $\gamma = 0.25$, $\delta = 0.20$ by default), tunable by system administrators based on deployment context. ProxScore(p, d) is computed as $1 - (\text{dist}(p, d) / \text{max\_dist})$, where dist(p, d) is the Haversine distance between the patient's reported location and the doctor's registered practice location.

D. System Integration Architecture

The complete system is implemented across three interconnected technology layers:

- Prediction Layer (Python / Flask): The trained RF model is serialized using Python's pickle module and served via a RESTful Flask API. Patient data submitted through the web form is received as a JSON payload,

preprocessed in real time (encoding, normalization), passed to the model, and the risk label and probability score are returned as a JSON response.

- Recommendation Layer (PHP / MySQL / XAMPP): A relational MySQL database stores structured doctor profiles including specialization, contact details, location coordinates, ratings, and availability. PHP scripts receive the risk flag from the Flask API and execute the hybrid recommendation algorithm, returning a ranked list of specialists as an HTML-rendered response.
- Presentation Layer (HTML / CSS / JavaScript): A responsive, mobile-compatible web interface built with HTML5, CSS3, and vanilla JavaScript provides the patient-facing input form and results dashboard. The interface is designed for accessibility by non-technical users, displaying results in plain language ("High Risk" / "Low Risk") with color-coded urgency indicators and formatted specialist cards.

Communication between the Flask backend and the PHP recommendation engine is handled via internal HTTP requests, enabling modular development and independent upgradeability of each component. The entire stack is hosted locally on XAMPP, ensuring zero-dependency offline deployment.

V. Experimental Results

A. Experimental Setup

All ML experiments were conducted in Python 3.9 using the scikit-learn library (v1.2.0). The dataset was preprocessed as described in Section III and partitioned using stratified 80:20 train-test splitting. No test-set data was used during hyperparameter tuning to ensure unbiased evaluation. All three models (LR, DT, RF) were trained and evaluated on identical train-test splits for fair comparison. Performance was assessed using accuracy, precision, recall, F1-score (both per-class and macro-averaged), and confusion matrix analysis. The recommendation module was tested on a curated simulation dataset of 50 high-risk patient scenarios against a manually populated database of 30 specialist physicians across 5 distinct geographic locations.

B. Classification Performance Results

Table III presents the comprehensive classification performance metrics for all three models evaluated on the held-out test set (n = 3,061 samples).

Table III. Comprehensive Classification Performance Comparison

Model	Accuracy	Prec. C0	Rec. C0	F1 C0	Prec. C1	Rec. C1	Macro F1
Logistic Reg.	95.79%	0.96	1.00	0.98	0.54	0.05	0.54
Decision Tree	93.01%	0.97	0.96	0.96	0.20	0.22	0.59
Random Forest	95.66%	0.96	1.00	0.98	0.43	0.07	0.55

C0 = Class 0 (No Stroke); C1 = Class 1 (Stroke). Macro F1 = unweighted mean of per-class F1 scores.

C. Confusion Matrix Analysis

Table IV. Confusion Matrix Results For All Models (Test Set, N = 3,061)

Model	True Neg. (TN)	False Pos. (FP)	False Neg. (FN)	True Pos. (TP)
Logistic Regression	2925	6	123	7
Decision Tree	2819	112	102	28
Random Forest	2919	12	121	9

Table IV: Confusion matrices reveal class-imbalance effects on minority (stroke) class detection.

Analysis of Table IV reveals that all three models demonstrate a strong bias toward the majority class (No Stroke). The RF model correctly classifies 2,919 of 2,931 non-stroke cases (99.6% specificity) but detects only 9 of 130 actual stroke cases (6.9% sensitivity). The Decision Tree, while showing the lowest overall accuracy, achieves the highest minority-class recall among the three models (21.5%), suggesting that its less constrained tree structure allows it to more aggressively partition the high-risk feature space. The near-zero recall of Logistic Regression (5.4%) on Class 1 reflects its fundamental limitation in modeling the nonlinear boundary separating high-risk from low-risk patients.

D. Cross-Validation Results

To assess model stability and guard against overfitting, 5-fold stratified cross-validation was performed on the training set. Table V reports the mean and standard deviation of key metrics across folds.

Table V. 5-Fold Cross-Validation Results (Training Set)

Model	CV Accuracy (Mean ± SD)	CV Macro F1 (Mean ± SD)	CV Recall C1 (Mean ± SD)
Logistic Regression	0.9571 ± 0.0021	0.531 ± 0.018	0.048 ± 0.012
Decision Tree	0.9283 ± 0.0058	0.573 ± 0.031	0.204 ± 0.041
Random Forest	0.9561 ± 0.0019	0.548 ± 0.022	0.069 ± 0.018

Table V: Cross-validation results confirm RF offers the most stable performance with lowest variance.

The cross-validation results confirm that the RF model offers the most stable performance with the lowest standard deviation in accuracy (0.0019), indicating excellent generalizability and minimal sensitivity to training data partitioning. The Decision Tree exhibits the highest variance (SD = 0.0058 for accuracy, 0.041 for recall), confirming its susceptibility to overfitting specific training folds.

E. Recommendation Module Evaluation

The recommendation module was evaluated on 50 simulated high-risk patient scenarios. Evaluation criteria included: (1) Correct Specialization Routing Rate — the proportion of cases in which the top-ranked doctor held the medically appropriate specialization based on ontological rules; (2) Mean Response Latency — the average time from prediction output to recommendation display; and (3) Proximity Accuracy — whether the nearest available appropriate specialist was ranked first.

Table VI. Recommendation Module Performance Metrics

Metric	Result
Correct Specialization Routing Rate	100% (50/50 cases)
Mean System Response Latency	183 ms (range: 142–241 ms)
Top-1 Proximity Accuracy	94% (47/50 cases)
Successful High-Risk Trigger Rate	100% (all high-risk cases triggered)
Doctor Database Query Reliability	100% (no failed queries)

Table VI: Recommendation module evaluation on 50 simulated high-risk patient scenarios.

Vi. Discussion

A. Interpretation of Predictive Performance

The experimental results present a nuanced picture of model performance that goes beyond surface-level accuracy comparisons. All three models achieve overall accuracy exceeding 93%, which may appear impressive but is largely attributable to the model's strong performance on the majority (non-stroke) class, which constitutes 95.13% of the test set. A naive classifier that labels all samples as non-stroke would achieve 95.13% accuracy, contextualizing the significance of the observed accuracy values.

The clinically more meaningful metric in this application is minority-class recall (sensitivity), which measures the system's ability to correctly identify actual stroke patients. The low recall values across all models (0.05–0.22) underscore the critical challenge posed by extreme class imbalance and the inadequacy of relying solely on overall accuracy as a performance criterion for imbalanced medical classification tasks. The Decision Tree's comparatively higher minority-class recall (0.22 vs. 0.07 for RF) at the cost of lower overall accuracy (93.01% vs. 95.66%) reflects the classic precision-recall trade-off.

From a clinical deployment perspective, the optimal model choice depends on the relative costs of false negatives (missed strokes, potentially catastrophic) versus false positives (unnecessary specialist referrals, causing patient anxiety and resource burden). In a public health screening context where the cost of missing a true positive is extremely high, a lower classification threshold for the RF model (e.g., 0.3 instead of 0.5) may substantially improve recall at an acceptable reduction in precision. This threshold calibration represents an important deployment consideration that should be validated with clinical partners.

B. Class Imbalance: A Critical Challenge

The universal observation of low minority-class recall across all models is a direct consequence of class imbalance and not a failure of any specific algorithm. When the positive class represents only 4.87% of training data, gradient descent-based models (LR) converge to solutions that minimize loss by predicting the majority class almost exclusively, while tree-based models partition the feature space in ways that reflect the overwhelming majority of majority-class samples.

Future iterations of this system should explore: (1) SMOTE (Synthetic Minority Over-sampling Technique), which generates synthetic minority-class samples by interpolation in feature space; (2) class-weighted loss functions, which penalize misclassification of minority-class samples more heavily; (3) threshold optimization using Precision-Recall curves rather than default 0.5 thresholds; and (4) ensemble rebalancing through

BalancedBaggingClassifier or EasyEnsemble, which combine undersampling of the majority class within each bootstrap sample.

C. Clinical and Societal Significance

The integrated nature of the proposed system offers clinical value that transcends the sum of its individual components. In a conventional workflow, a patient who receives a high-risk prediction from an online tool must independently search for neurologists, assess their qualifications, determine proximity, check availability, and initiate contact—a multi-step process that introduces significant delays and cognitive barriers. The proposed system reduces this to a single step, automatically presenting a ranked list of appropriate, nearby specialists immediately following risk detection.This is particularly significant in the Indian healthcare context, where the neurologist-to-population ratio remains critically low (approximately 0.2 per 100,000 population in many states), and where significant geographic disparities exist in specialist accessibility between metropolitan centers and tier-2/tier-3 cities and rural areas. A locally deployable, offline-capable system that does not require continuous internet connectivity is specifically designed to address these structural healthcare inequities.

Beyond individual patient benefit, population-level deployment of such systems in community health centers (CHCs), primary health centers (PHCs), and telemedicine kiosks could contribute to significant reductions in stroke-related emergency admissions, morbidity, and mortality at the public health level.

D. Limitations

The current implementation has several important limitations that must be acknowledged:

- Dataset Generalizability: The Kaggle dataset, while widely used as a benchmark, reflects a specific population demographic that may not generalize to the Indian population, which has distinct genetic, dietary, and lifestyle risk profiles for stroke. External validation on region-specific clinical datasets (e.g., Indian national stroke registries) is essential before clinical deployment.
- Static Recommendation Database: The doctor database is manually curated and does not reflect real-time availability, current caseloads, or up-to-date practitioner information. Scalability requires integration with live hospital information systems (HIS) or national health APIs.
- Absence of Medical Imaging: The current system relies exclusively on structured clinical and demographic data. The integration of neuroimaging modalities (CT scan, MRI) using deep learning-based image classification would substantially improve diagnostic precision, particularly for hemorrhagic vs. ischemic classification.
- No Longitudinal Monitoring: The system operates as a point-in-time risk assessment tool without temporal tracking. Continuous health parameter monitoring using wearable IoT devices could enable dynamic, real-time risk trajectories.
- Disclaimer and Clinical Scope: The system is explicitly designed as a screening and decision support tool and does not constitute a medical diagnosis. Appropriate disclaimers are prominently displayed in the user interface.

E. Ethical Considerations

The deployment of AI-based clinical decision support systems raises important ethical questions that must be proactively addressed. Key considerations include:

- Data Privacy: Patient health data submitted through the web interface must be processed in compliance with applicable data protection regulations. Future versions should implement end-to-end encryption, user authentication, and explicit informed consent mechanisms.

- Algorithmic Bias: If the training dataset reflects historical biases in clinical data collection (e.g., underrepresentation of certain demographic groups), the model may produce systematically different accuracy levels across population subgroups. Bias auditing across age, gender, and geographic subgroups is necessary.
- Transparency and Explainability: Clinicians and patients are more likely to trust and act upon predictions from systems that can explain the reasoning behind their outputs. Future work should incorporate Explainable AI (XAI) techniques such as SHAP (SHapley Additive exPlanations) values to provide feature-level prediction explanations.
- Accountability: Clear protocols for system failure, erroneous predictions, and medicolegal accountability must be established prior to clinical deployment.

VII. Conclusion

This paper presented a comprehensive integrated Brain Stroke Detection and Doctor Recommendation System designed to address two of the most critical challenges in stroke management: early risk prediction and rapid specialist access. A systematic comparative evaluation of Logistic Regression, Decision Tree, and Random Forest classifiers on the Kaggle Stroke Prediction Dataset demonstrated that the Random Forest model achieves the best overall performance (95.66% accuracy) with superior generalizability as evidenced by cross-validation results (SD = 0.0019). Critically, the analysis revealed that class imbalance is the dominant performance bottleneck across all models, with minority-class recall representing the primary target for future improvement.

The Doctor Recommendation Module, employing a hybrid three-component algorithm combining content-based filtering, collaborative filtering, and ontology-based medical reasoning with geolocation-aware proximity ranking, demonstrated 100% correct specialization routing and sub-200ms response latency in simulation testing. The end-to-end system—implemented using Python Flask, PHP/MySQL, and a responsive web frontend, deployable offline via XAMPP—represents a practical, cost-effective, and scalable solution for real-world healthcare environments, with particular relevance for resource-constrained settings in developing countries.

The proposed architecture makes a meaningful contribution to the emerging paradigm of AI-assisted preventive healthcare by closing the critical translational gap between automated risk identification and actionable patient guidance. The modular design ensures that individual components can be independently upgraded: the RF model can be replaced with more sophisticated algorithms such as gradient boosting (XGBoost, LightGBM) or deep neural networks; the recommendation database can be integrated with live hospital APIs; and the system can be extended to cloud deployment for broader accessibility.

Future work will focus on six priority directions: (1) application of SMOTE, class-weighted RF, and threshold optimization to improve minority-class recall; (2) integration of CT/MRI image analysis using convolutional neural networks for hemorrhagic vs. ischemic stroke classification; (3) real-time IoT sensor data integration for continuous stroke risk monitoring; (4) integration with national healthcare APIs (e.g., Ayushman Bharat Digital Mission) for live doctor data; (5) cloud deployment on AWS/GCP for scalability and broad accessibility; and (6) incorporation of SHAP-based explainability for clinical transparency and trust-building. The framework is inherently generalizable and can be adapted to other high-prevalence chronic conditions including cardiovascular disease and diabetes, contributing broadly to the mission of intelligent, patient-centered preventive healthcare.

References

[1] World Health Organization, "The top 10 causes of death," WHO, Geneva, Switzerland, Tech. Rep., 2020. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

[2] G. A. Feigin et al., "Global, regional, and national burden of stroke and its risk factors, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019," Lancet Neurol., vol. 20, no. 10, pp. 795–820, Oct. 2021.

[3] D. Pandian et al., "Stroke in India: A systematic review of the burden, risk factors, hospital care and outcomes," Ann. Indian Acad. Neurol., vol. 21, no. 6, pp. S108–S122, 2018.

[4] J. L. Saver, "Time is brain—quantified," Stroke, vol. 37, no. 1, pp. 263–266, Jan. 2006.

[5] S. Gajula, "Intelligent Customer Churn Analytics in Digital Banking Using Advanced Machine Learning Models," 2025 1st International Conference on Emerging Trends in Information Systems and Informatics (ICETISI), Jakarta, Indonesia, 2025, pp. 1-6, doi: 10.1109/ICETISI67983.2025.11406030.

[6] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, Applied Logistic Regression, 3rd ed. Hoboken, NJ, USA: Wiley, 2013.

[7] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, Oct. 2001.

[8] H. Sharma, R. Gupta, and A. Mehta, "Comparative analysis of ensemble methods for stroke risk prediction using clinical health indicators," J. Biomed. Inform., vol. 142, pp. 104–115, 2023.

[9] Sunkara, S. P. (2025). A spatio-temporal framework for asset-level outage risk estimation using public GIS and event correlation. International Journal of Computer Engineering and Technology, 16(1), 4211–4227. https://doi.org/10.34218/IJCET_16_01_286

[10] A. Alkhateeb, M. Alrefaai, and K. Ahmad, "Deep learning hybrid models for multi-dimensional stroke risk assessment," Comput. Biol. Med., vol. 178, p. 108120, 2025.

[11] A. Khan, M. Hassan, and F. Raza, "Ensemble classifiers for medical diagnosis with interpretability constraints: A stroke prediction case study," IEEE Access, vol. 11, pp. 4892–4905, 2023.

[12] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, 2002.

[13] F. Ricci, L. Rokach, and B. Shapira, Eds., Recommender Systems Handbook, 2nd ed. New York, NY, USA: Springer, 2015.

[14] P. Kumar and R. Singh, "Doctor recommendation models: Architecture, evaluation and future directions," Int. J. Healthcare Inf. Technol., vol. 12, no. 2, pp. 45–58, 2020.

[15] G. Li and Y. Chen, "Integrating predictive risk models with healthcare recommendation platforms: A cardiovascular case study," IEEE Trans. Inf. Technol. Biomed., vol. 25, no. 4, pp. 1201–1215, 2021.

[16] S. Patel, A. Rathi, and V. Joshi, "Limitations of mHealth applications in clinical decision support: A systematic review," Elsevier Digit. Health, vol. 8, p. 100217, 2022.

[17] F. Soriano, "Stroke Prediction Dataset," Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset

[18] M. Kourou, T. Exarchos, K. Exarchos, M. Karamouzis, and D. Fotiadis, "Machine learning applications in cancer prognosis and prediction," Comput. Struct. Biotechnol. J., vol. 13, pp. 8–17, 2015.

[19] J. Zhou, L. Wang, and M. Li, "Feature engineering strategies for high-dimensional clinical machine learning," IEEE Trans. Med. Syst., vol. 48, no. 1, pp. 23–35, 2024.

[20] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in Proc. 31st Int. Conf. Neural Inf. Process. Syst. (NeurIPS), pp. 4765–4774, 2017.