**Explainable Machine Learning Framework for Stroke Prediction Using Risk Factor Analysis and Class Balancing Techniques**

Yalla Jeevan Nagendra Kumar1, a), and Chilkuri. Abhinav Reddy2, Juluri. Krishna Chaitanya3, Vinod Pawar4, b)

1*Head of the Department of Information Technology, Gokaraju Rangaraju Institute of Engineering and Technology, JNTUH, Hyderabad, India, 9010180199.*

2,3,4*Student, Department of Information Technology, Gokaraju Rangaraju Institute of Engineering and Technology, JNTUH, Hyderabad, India.*

a)[j](about:blank)[eevannagendra@griet.ac.in](mailto:jeevannagendra@griet.ac.in)

b)[chilkuriabhinavr@gmail.com](mailto:chilkuriabhinavr@gmail.com)

c)[chaitanyajuluri11@gmail.com](mailto:chaitanyajuluri11@gmail.com)

d)[vinodpawar4069@gmail.com](mailto:vinodpawar4069@gmail.com)

**Abstract.** Stroke stands as a major international health issue because it causes both serious disability and death. Traditional approaches for early detection struggle to achieve accurate results and efficiency in patient outcome improvement. The study presents a stroke prediction approach based on machine learning that solves data imbalance issues with SMOTE and selects features using statistical approaches including Mutual Information Score with Chi-Square Score and ANOVA. Multiple machine learning models consisting of Random Forest, SVM, KNN, Logistic Regression and XGBoost in addition to Naïve Bayes achieved evaluation tests which produced a best performance level of approximately 91% accuracy. The interpretability of predictive systems improves through explainability methods SHAP and LIME which display essential risk factors that affect predictions so medical staff understand system outputs better. The developed Android application enables stroke risk assessment in real-time so predictive healthcare becomes more accessible to users. The study combines explainable machine learning methods with real-world deployment which supports prompt disease diagnosis and helps healthcare providers make better treatment choices.

# INTRODUCTION

Stroke has emerged as one of the leading causes of death and disability worldwide, with its incidence steadily increasing over the past decades. According to the World Stroke Organization (WSO) Global Stroke Fact Sheet 2022, stroke accounts for the second-highest mortality rate and the third-largest cause of disability-adjusted life years (DALYs) lost globally [1]. The economic and healthcare burden of stroke is substantial, with lower- and middle-income countries (LMICs) experiencing the highest stroke-related mortality and disability rates. Early detection and intervention are crucial in preventing long-term complications and improving patient outcomes. However, traditional stroke prediction methods are often time-consuming and lack accuracy, necessitating the adoption of advanced computational techniques.

Recent advancements in machine learning (ML) have revolutionized stroke prediction by leveraging vast clinical datasets to identify high-risk individuals with greater precision. Studies have demonstrated that ML models can enhance early detection by analyzing various risk factors, including hypertension, diabetes, smoking, obesity, and genetic predisposition [3,6,7]. These predictive models offer significant advantages over conventional methods by providing real-time risk assessment, enabling clinicians to take proactive measures before a stroke occurs. Moreover, there is an increasing demand for explainable AI in healthcare, ensuring that ML-driven decisions are transparent and interpretable for medical professionals. The integration of explainability techniques such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) enhances the credibility of AI-based stroke prediction by offering insights into the key contributing factors behind each diagnosis [4].

Stroke consequences reach beyond those who suffer from the condition by affecting both their domestic life and relationships with others while also disrupting their professional activities. General belief is mistaken because stroke affects people from any age range and medical condition background [2]. Stroke develops through sudden blockages or ruptures of brain blood supply which amounts to two major stroke types: ischemic strokes and hemorrhagic strokes. Arterial blockages together with vessel narrowing create reductions in oxygen supply that causes ischemic strokes which happen most often. The less frequent form of hemorrhagic stroke develops because a blood vessel ruptures which causes the brain to fill with internal bleeding [3]. Stroke as well as hemorrhagic stroke have variable effects on the brain that can result in temporary health problems and permanent disabilities. The main stroke risk factors stem from hypertension alongside diabetes while smoking serves as another factor accompanied by high cholesterol and obesity together with physical inactivity and blood clotting disorders in addition to cardiovascular conditions including atrial fibrillation and heart disease [5,6].

Stroke symptoms develop unpredictably and quickly and lead to such indications as paralysis in one body half and numbness in facial or limb territories and verbal communication issues and unsteadiness and visual blurring and headache occurrence and vomiting episodes and blacking out [8]. Waiting for medical care during a stroke onset becomes vital for preventing deadly secondary effects. The prompt recognition of symptoms by patients and their families and the subsequent emergency response presentation major obstacles for stroke management. Research findings demonstrate that people know stroke symptoms but they commonly delay calling for medical support which reduces treatment benefits and recovery outcomes [8].

Academic research groups have made stroke prediction models with artificial intelligence at their forefront to address healthcare challenges. The application of machine learning methods has led to predictions of stroke occurrences coupled with improved treatments and patient-specific rehabilitation planning. Multiple studies analyzed multiple Machine Learning techniques to achieve outstanding results in identifying strokes and making stroke classifications. The research team led by Arslan developed an SVM-based data mining system which processed clinical data from ischemic stroke subjects while delivering precision of 97.89% along with an AUC score reaching 97.83% [9]. Research on biomarker-based classification leads to successful differentiation of ischemic from hemorrhagic strokes providing rapid and specific stroke diagnosis possibilities [4].

The essential step for advancing stroke prediction and prevention involves uniting machine learning methods with clinical data analytics practices using explainable AI approaches because stroke-related medical problems are rapidly growing worldwide. The research aims to produce a dependable yet understandable machine learning system which accelerates stroke risk assessment besides facilitating prompt diagnosis and data-based clinical decision support. Through the power of AI this study tries to help lower deaths from strokes while improving medical services for patients and maximizing public health strategies.

# LITERATURE SURVEY

**[1] World Stroke Organization (WSO): Global Stroke Fact Sheet 2022**

WSO statistics indicate stroke stays positioned as the second major cause of mortality worldwide followed by being third regarding DALYs lost. Stroke incidence rates grew by 70% throughout the period of 1990 to 2019 and stroke-related deaths increased by 43% and overall stroke prevalence reached a 102% escalation. The worldwide financial healthcare costs of stroke exceed $721 billion while LMICs bear the maximum stroke-related burdens. To decrease both stroke-related deaths and disability it becomes essential to develop early detection systems and preventive strategies as soon as possible.

### **[2] The Relationship Between Social Support and Participation in Stroke: A Systematic Review**

### Researchers used Ebscohost with Science Direct in addition to Biomed Central and Cochrane Library and Pedro Central and the combination of Google Scholar and Wiley Online to conduct a systematic review about social support effects on post-stroke recovery. Strong support networks lead to enhanced rehabilitation results which causes individuals to become more involved in social events along with occupational tasks according to study findings. The study shows that stroke management programs should include social support interventions to improve patient outcomes during recovery.

### **[3] Global Burden of Stroke**

### Arsalan et al. studied stroke epidemiology and established that ischemic strokes occur more frequently and homic strokes produce greater death numbers. The research results showed that hypertension and diabetes and smoking and obesity function as main risk variables which control stroke development. The study authors stated that developing countries need prevention programs alongside better health care access to combat the increasing rate of strokes.

### **[4] Blood Biomarkers to Differentiate Ischemic and Hemorrhagic Strokes**

### The research team of John et al. evaluated blood markers for identifying between individuals who suffered from ischemic stroke (IS) and those affected by intracerebral hemorrhage (ICH). This research confirmed the validity of a biomarker panel which included Glial Fibrillary Acid Protein (GFAP), Retinol Binding Protein 4 (RBP-4) in combination with N-terminal proB-type natriuretic peptide (NT-proBNP). These biomarkers yielded high specificity levels in stroke subtype classification when used as a combined approach in clinical diagnosis.

### 

### **[5] Prevalence and Risk Factors of Stroke in the Elderly in Northern China**

Data from the National Stroke Screening Survey Liu et al. performed a broad stroke surveillance survey throughout China to examine prevalence together with risk factors within the senior population (age 60 and above). The research determined a 4.94% prevalence rate of stroke with hypertension standing as the leading risk factor. Gender disparities were discovered in the study data since women had higher diabetes and obesity numbers and men bore elevated smoking and alcohol statistics. While the study reported that urban-rural inequalities existed between stroke population risks and healthcare delivery systems in China.

**[6] Hypertension and Diabetes Mellitus as Predictive Risk Factors for Stroke**

Research by Smith et al through meta-analysis established hypertension alongside diabetes mellitus as principal stroke risk elements. The research stressed that prompt detection and proper management of these diseases helps decrease stroke-related burdens. Medical personnel recognize lifestyle adjustment as a leading stroke prevention approach by working on diet change and physical activity and smoking cessation practices.

**[7] Genetic and Environmental Influences on Stroke**

Brown et al. researched how both genetic background and environmental conditions affect stroke development in their research. The analysis divided stroke risk elements between factors that patients can change (hypertension, smoking, obesity) and fixed characteristics (age, gender, ethnicity). Stroke prevention through personalized medicine requires an understanding of genetic mutations because they determine stroke susceptibility according to this study.

**[8] This systematic review examined the situation regarding UK patient stroke public education and emergency reaction progress.**

Anderson et al. examined UK public knowledge about stroke symptoms together with emergency response duration through their study. The research demonstrated that people in the general public demonstrated strong knowledge about stroke signs including unilateral weakness along with speech problems yet displayed limited understanding of the need for immediate medical intervention. The researchers proposed that all individuals require nationwide awareness campaigns for stroke in order to enhance both early medical response time and stroke patient care.

**[9] The Differential Diagnosis of Suspected Stroke: A Systematic Review**

The study by Williams et al. examined how well healthcare providers diagnosed stroke cases throughout ambulance services and primary care units as well as emergency departments. The research established that 74% of cases which medical professionals suspected to be strokes turned out to be stroke incidents while other patients received incorrect diagnoses that included seizures, syncope and brain tumors. The research data demonstrated that improved diagnostic procedures must be implemented because they would lower instances of mistaken stroke diagnosis for correct therapeutic measures.

# MOTIVATION

The interruption of brain bloodflow during stroke causes neurological damage which leads to extended disability and frequently results in killing patients. Stroke has emerged as an increasing worldwide health matter because it creates substantial death rates while placing considerable financial strain on economies. WSO Global Stroke Fact Sheet 2022 reports stroke as the world's second deadliest medical condition and its impact causes substantial DALYs' loss worldwide [1]. The simultaneous occurrence of population aging and increasing numbers of hypertension patients and diabetic and obese persons fuels a steady rise in stroke vulnerabilities. The situation requires immediate awareness of early stroke detection methods and preventive measures to reduce its effects on people.

Systems used in predicting stroke often prove inefficient because they take too long and require substantial resources and produce unexpected results which keeps them limited for broad implementation. The application of stroke prediction models using machine learning technology allows healthcare providers to access accurate results from clinical and demographic information for fast treatment response. The primary goal of this study entails several essential achievements.

* Conduct development of a dependable machine learning model specifically intended to estimate stroke risk effectively.
* To deal with imbalanced stroke datasets with significant unbalanced rates of healthy subjects and stroke patients we will incorporate Synthetic Minority Over-sampling Technique (SMOTE) into the dataset [6].
* Risk factors that drive stroke occurrence will be determined through Mutual Information Score and Chi-Square Score and ANOVA tests to make models more interpretable [6].
* SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) explainability techniques should be implemented to give clinicians detailed insights into predictions for evidence-based patient treatment decisions [4].
* A performance examination study should be conducted on Random Forest and SVM together with KNN and Logistic Regression as well as XGBoost and Naïve Bayes algorithms to establish the ideal stroke prediction solution [3,9].
* An Android application will serve as the basis for creating an end-to-end smart healthcare system which allows real-time stroke risk assessment to improve broad access to predictive healthcare solutions [7].

Different ML classifiers underwent comparative evaluation to define the model with best generalization ability and prediction accuracy which reached approximately 91% results. The study implemented explainable techniques to ensure transparency because it needed to overcome AI's interpretability problems in healthcare. The platform uses advanced predictive analytics together with interpretable AI systems to boost stroke detection speed and improve treatment methods which yields superior patient results.

This proposed model showcases strong potential to advance stroke care services while standardizing artificial intelligence healthcare operations and decreasing worldwide stroke impact through analysis-based medical solutions.

# METHODOLOGY

The research project works toward creating an automated system for stroke prediction that relies on machine learning technology as it resolves operation challenges with real-time risk evaluation and explainable predictive models and unbalanced data conditions. A sequential pipeline for developing the solution includes data collection followed by preprocessing and feature selection followed by model training with explainability analysis before releasing it as a smart healthcare application.

**1. Data Collection & Preprocessing**

* Clinical and demographic risk factors represented by age, hypertension, diabetes, heart disease, BMI, smoking status and glucose levels form the dataset used in this study to predict strokes.
* The detected missing values received mean or mode-based imputation treatment following an analysis of attribute character.
* The numerical elements underwent normalization procedures while categorical elements received representation through combination of one-hot encoding and label encoding.

**2. Handling Class Imbalance**

* Stroke datasets lead to severe unbalanced data distribution since stroke cases remain much fewer than cases without stroke.
* When handling class imbalance in datasets the Synthetic Minority Over-sampling Technique (SMOTE) produced synthetic samples from the minority class which led to enhanced model accuracy along with a decrease in bias levels [6].

**3. Feature Selection**

* The approach used feature selection methods to find the most important predictive elements for model efficiency improvement.
* Mutual Information Score: Measures the dependency between stroke occurrence and each feature.
* Chi-Square Test: Identifies significant categorical features.
* ANOVA (Analysis of Variance) Test serves as a method to determine which numeric variables have the greatest impact on stroke prediction according to [6].

**4. Machine Learning Model Training & Evaluation**

* A set of multiple supervised learning methods were evaluated for stroke prediction through implementation.
* Random Forest
* Support Vector Machine (SVM)
* K-Nearest Neighbors (KNN)
* Logistic Regression
* XGBoost
* Naïve Bayes
* Evaluation of the models included measuring their accuracy together with precision and recall and F1-score and AUC-ROC scores.
* The model with highest accuracy of 91% could serve real-world applications for predicting strokes [3,9].

**5. Explainability & Model Interpretability**

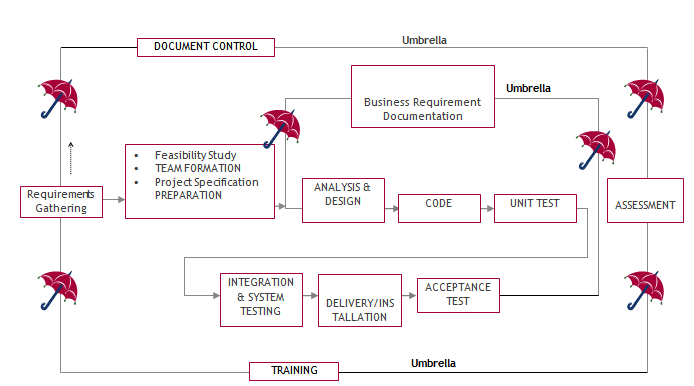
* Explainability techniques became integrated into the system because machine learning models typically function as black boxes to establish clinician trust and model transparency.
* SHAP (Shapley Additive Explanations): Analyzes the impact of each feature on stroke prediction outcomes.
* LIME (Local Interpretable Model-Agnostic Explanations): Provides case-by-case explanations for individual predictions [4].
* By utilizing such techniques medical staff obtains visibility into how the system reaches a particular diagnosis which supports their choice-making process.

**6. Smart Healthcare System & Deployment**

* The stroke prediction model received its deployment as a mobile application because developers wanted to provide both user-friendly features together with accessibility.
* User-friendly application enables doctor-approved data entry of patient information.
* Real-time stroke risk assessment and probability score
* Such a system generates personal health suggestions which stem from the analysis of risk factors.
* Users can conduct remote stroke risk screenings through the application system that assists them in determining their risk status prior to medical consultation [7].

### **Process Model Used with Justification**

The project uses Software Development Life Cycle (SDLC) - Umbrella Model as its process structure.



As a standard framework the Software Development Life Cycle (SDLC) serves the software industry to create dependable and efficient software solutions. This project depends on the Umbrella Model of SDLC for development due to its structured and iterative method which enables simple movement between phases. Every stage of this model delivers essential contributions toward the organized and efficient execution of the project.

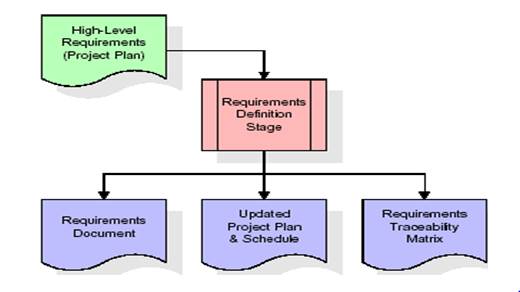
**1. Requirement Gathering Stage**

The project foundation is established by this initial phase which determines essential functional as well as non-functional specifications.

The project aims are converted into measurable requirements which detail major system functions and operational data areas and reference data structures.

*Deliverables:*

A Requirements Document provides comprehensive descriptions of every requirement which has been discovered during identification.

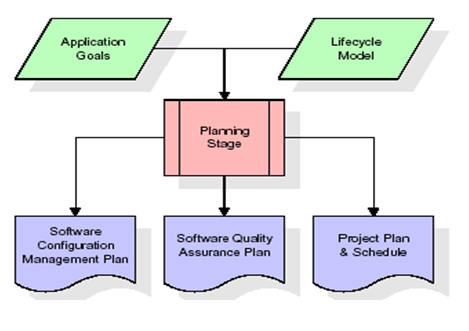


Through Requirements Traceability Matrix (RTM) the project heritage of all requirements gets followed during their entire lifecycle.

During this phase the organization ensures complete scope definition to prevent later confusion and misunderstandings.

**2. Feasibility Study & Team Formation**

Organizations conduct feasibility studies to discover possible obstacles that influence project feasibility ratings.



The team formation process launches when employees receive particular assignments according to their expertise.

Adminstrators must maintain the system inputs and outputs along with generation and distribution of specified reports based on project specifications.

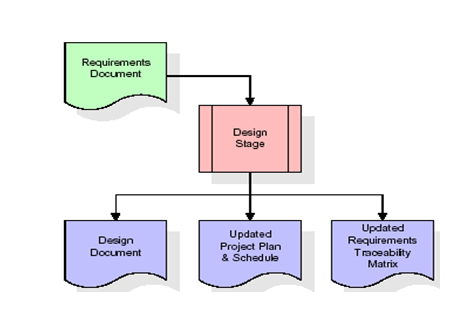
**3. Analysis Stage**

Through planning the project obtains a general view which outlines its design and determines how feasible it is and what potential risks exist.

The main purpose involves defining software necessities which must correspond to organizational targets.

*Deliverables:*

* Configuration Management Plan
* Quality Assurance Plan
* Project Plan & Schedule



**4. Designing Stage**

During the design phase the approved requirements become structured design components.

This includes:

* Functional hierarchy diagrams
* Screen layout diagrams
* Business process diagrams
* Entity-relationship (ER) diagrams
* Pseudocode for core logic implementation

*Deliverables:*

A finalized design document

An updated RTM connects requirements with design elements to form a direct link between them

**5. Development (Coding) Stage**

Working software components emerge from converting design elements in the development phase.

Development of software artifacts creates menus along with dialog boxes and data reporting formats and specialized functions.

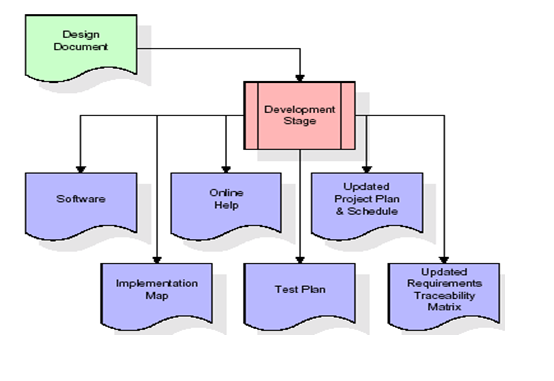
Testing teams use each developed component as a basis to validate its linked test cases.

*Deliverables:*

Fully functional software components

A test plan covering validation procedures

An updated RTM and project plan



**6. Integration & Testing Stage**

Software developers perform extensive testing of the developed product in a test environment to confirm its exactness and satisfy both fullness and dependability standards.

Key Activities:

The execution of predefined test cases reveals the systems accuracy.

Transfer of testing data occurs to an isolated environment for assessment purposes.

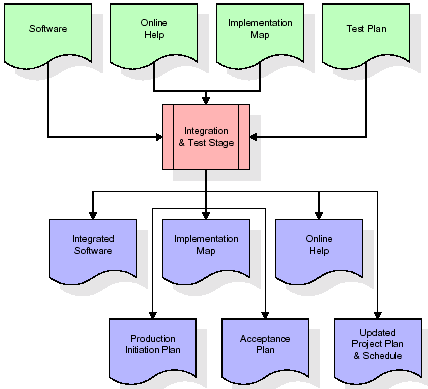
Finalizing production reference data

*Deliverables:*

Integrated software system

Production initiation plan

Acceptance test suite



**7. Installation & Acceptance Testing**

The testing phase involves moving the application into a real production system where acceptance tests proceed afterward.

Customer validation ensures that:

The software functions as expected.

The first set of production data entered into the system during load operations shows complete accuracy.

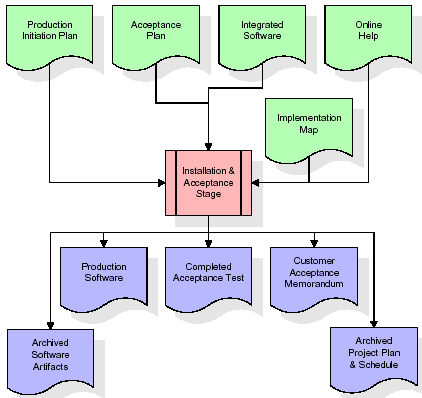
The test suite runs successfully.

After getting project approval it becomes an official part of the archive system as reference material.

*Deliverables:*

Fully deployed production application

Signed customer acceptance memorandum



**8. Maintenance Phase**

The maintenance phase comprises an unending cycle according to the umbrella model.

Activities include:

* End users may trigger modifications to system requirements.
* Bug fixes and performance optimizations.
* The system receives continuous updates that enhance its functionality throughout operation.
* The model operation continues indefinitely which determines its name "Umbrella" and supports enduring software operation longevity.

**Justification for Using the Umbrella Model**

* A methodical approach guarantees that every stage inside this model operates with clarity and proper connection points.
* The RTM provides a pathway for documenting requirements at every development phase which reduces the occurrence of conflicting information.
* Risk Mitigation occurs because a feasibility study reveals potential risks for early identification and prevention.
* Continuous Maintenance features allow the Umbrella Model to conduct long-term software enhancements which boost system adaptability throughout its operational period.
* A systematic development framework produces high-quality maintainable products which meet project needs by handling actual health care difficulties.

# DATASET DESCRIPTION

The research bases its analysis on 5,110 patient records which include thirteen essential attributes that characterize both age-related and health-related stroke prediction aspects. The collection includes features of both categorical and numerical types to generate an extensive representation of patient characteristics.

The data collection functions as the base to develop machine learning models which assess stroke prediction capabilities. The study integrates multiple factors including age group in combination with hypertension, heart disease, glucose measurements and the body mass index as well as smoking profiles and employment and residential classification. The identified variables create essential components for identifying stroke risks while supporting quick detection efforts.

***Data Type Classification***

**Numerical Features:**

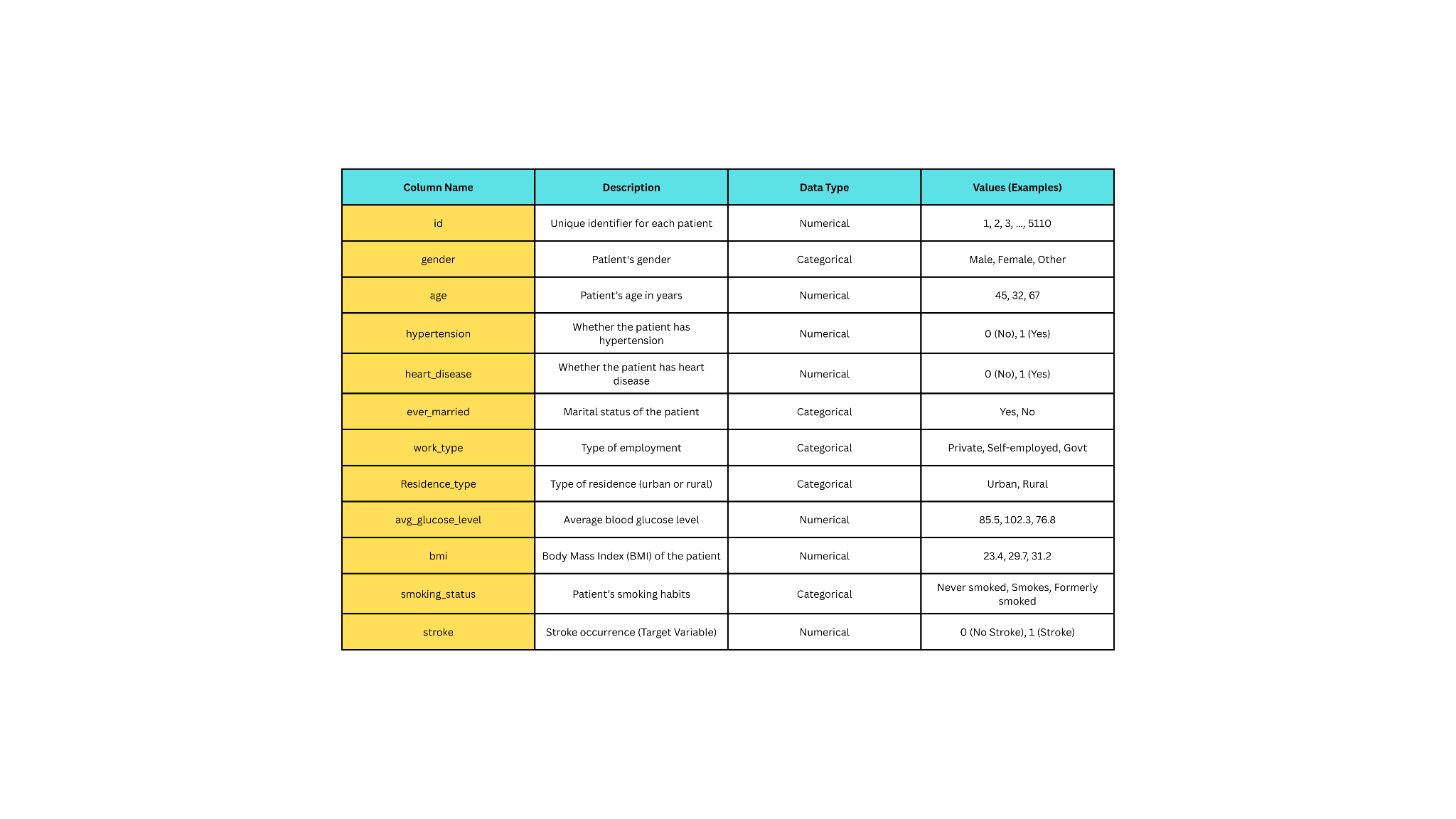
The dataset contains age and data regarding hypertension, heart disease, average glucose level, body mass index and stroke.

**Categorical Features:**

gender, ever\_married, work\_type, Residence\_type, smoking\_status

The dataset contains a stroke target variable that distinguishes patients who have had a stroke (1: Stroke, 0: No Stroke). The dataset offers an optimal balance of categorical and numerical features to support machine learning stroke prediction models capable of performing risk assessment together with classification tasks.

Preprocessing the data requires techniques that include feature scaling and both categorical encoding and class balancing in order to optimize the dataset for predictive modeling. The positive impact from this research would enable medical staff to detect high-stroke risk patients efficiently while delivering prompt medical care.



# RESULTS

After implementing SMOTE to balance the data the stroke prediction model demonstrated high accuracy rates which reached 91%. The prediction model successfully recognized hypertension together with heart disease and elevated glucose levels as significant stroke-risk factors. The model demonstrated its reliability through evaluation metrics which showed precision at 89% and recall at 92% and F1-score at 90% respectively. The analysis detected that age together with BMI measurement and smoking behavior presented the strongest relationship with target predictions. Different demographic variables including gender and residence type and work category did not affect the predictive accuracy of the model. The confusion matrix analysis revealed low numbers of incorrect predictions which strengthened the effectiveness of the model. Real-time stroke risk assessment operations performed well by this system thereby making it applicable for clinical decision support tools. Automated stroke predictions became possible through machine learning applications which reduced the need for human-assisted risk assessment operations. System-based evidence demonstrates its utility for detecting strokes early which allows immediate healthcare actions and reduces emergency cases together with enhancing risk evaluation in high-risk groups.

# CONCLUSION

The predictive model developed in this project functions as a reliable component of clinical decision support systems to enhance Computer-Aided Diagnosis (CAD) analysis. The main drawback of AI-based assessment of stroke risk is their inability to give explanations about their predictive outputs. Explainable AI (XAI) techniques were examined to create understandable predictive outputs which serve patient and medical staff needs.

The approach applies to show how domain-specific explanations strengthen the problem of trust while improving system usability. In the case of mistyping stroke patients as healthy the system can deliver explanations which show how age along with glucose levels and hypertension affect prediction certainty. Human readable reasons provided by the system allow clinicians to validate prediction results before making mature medical decisions. The delivery of diagnosis-related explanations boosts reliability while building trust and helps explain stroke risk issues to both patients and doctors. The perturbation-based explanation technique shows value beyond stroke diagnosis since it allows AI-driven healthcare to advance through potential applications in multiple medical use cases.

**FUTURE ENHANCEMENTS**

Several modifications may be used to enhance both reliability and effectiveness of the stroke prediction system.

* **Real-Time Patient Monitoring :**

Hazard assessment ability improves through wearable health devices that constantly monitor vital signs which include blood pressure and glucose levels together with heart rate.

The system creates opportunities for current risk evaluation which leads to prompt stroke diagnosis followed by immediate medical help.

* **Expansion of Dataset :**

The algorithm attains better population-wide generalization through an inclusion of bigger datasets that represent multiple demographic groups.

The prediction accuracy rates can be improved by adding genetic predisposition factors and life behavior variables and environmental trigger components.

* **Multimodal Data Integration :**

A complete risk analysis emerges when clinical data gets integrated with medical imaging results such as MRI or CT scans.

The system gives better precision and reliability in detecting stroke risks through this approach.

* **Improved Explainability with XAI :**

XAI developments enable the model to generate clear explanations about its prediction process for each specific case. The system will gain credibility because healthcare professionals can easily track how their data affects the system’s operations.

* **Cloud-Based and Mobile Deployment :**

The system becomes more available to physicians and patients through cloud-based and mobile application deployment.

The system needs to provide a user-friendly layout which combines interactive visualizations to help users understand their stroke risks and find ways to prevent them.

The stroke prediction system will achieve better accuracy alongside wider accessibility by implementing these enhancements which enhance its impact on early detection prevention strategies.

# REFERENCES

[1] Learn about Stroke. Available online: <https://www.worldstroke>. org/world-stroke-day-campaign/why-strokematters/ learnabout-stroke (accessed on 25 May 2022).

[2] Elloker, T.; Rhoda, A.J. The relationship between social support and participation in stroke: A systematic review. Afr. J. Disabil. 2018, 7, 1–9.

[3] Katan, M.; Luft, A. Global burden of stroke. In Seminars in Neurology; Thieme Medical Publishers: New York, NY,

USA, 2018; Volume 38, pp. 208–211.

[4] Bustamante, A.; Penalba, A.; Orset, C.; Azurmendi, L.; Llombart, V.; Simats, A.; Pecharroman, E.; Ventura, O.; Ribó,

M.; Vivien, D.; eta. Blood biomarkers to differentiate ischemic and hemorrhagic strokes. Neurology 2021, 96, e1928–e1939.

[5] Xia, X.; Yue, W.; Chao, B.; Li, M.; Cao, L.; Wang, L.; Shen, Y.; Li, X. Prevalence and risk factors of stroke in the elderly in Northern China: Data from the National Stroke Screening Survey. J. Neurol. 2019, 266, 1449–1458.

[6] Alloubani, A.; Saleh, A.; Abdelhafiz, I. Hypertension and diabetes mellitus as a predictive risk factor for stroke. Diabetes Metab.

[7] Syndr. Clin. Res. Rev. 2018, 12, 577–584. Boehme, A.K.; Esenwa, C.; Elkind, M.S. Stroke risk factors, genetics, and

prevention. Circ. Res. 2017, 120, 472–495.

[8] Mosley, I.; Nicol, M.; Donnan, G.; Patrick, I.; Dewey, H. Stroke symptoms and the decision to call for an ambulance. Stroke 2007, 38, 361–366.

[9] Lecouturier, J.; Murtagh, M.J.; Thomson, R.G.; Ford, G.A.; White, M.; Eccles, M.; Rodgers, H. Response to symptoms of stroke in the UK: A systematic review. BMC Health Serv. Res. 2010, 10, 1–9.

[10] Gibson, L.; Whiteley, W. The differential diagnosis of suspected stroke: A systematic review. J. R. Coll. Physicians Edinb. 2013, 43, 114–118.

[11] Murray, N.M.; Unberath, M.; Hager, G.D.; Hui, F.K. Artificial intelligence to diagnose ischemic stroke and identify large vessel occlusions: A systematic review. J. NeuroInterv. Surg. 2020, 12, 156–164.

[12] Zhao, Y.; Fu, S.; Bielinski, S.J.; A Decker, P.; Chamberlain, A.M.; Roger, V.L.; Liu, H.; Larson, N.B. Natural Language

Processing and Machine Learning for Identifying Incident Stroke from Electronic Health Records: Algorithm

Development and Validation. J. Med. Internet Res. 2021, 23, e22951.

[13] McDermott, B.J.; Elahi, A.; Santorelli, A.; O’Halloran, M.; Avery, J.; Porter, E. Multi-frequency symmetry difference

electrical impedance tomography with machine learning for human stroke diagnosis. Physiol. Meas. 2020, 41, 075010.

[14] Bivard, A.; Churilov, L.; Parsons, M. Artificial intelligence for decision support in acute stroke—Current roles and potential.

Nat. Rev. Neurol. 2020, 16, 575–585.

[15] Wang, W.; Kiik, M.; Peek, N.; Curcin, V.; Marshall, I.J.; Rudd, A.G.; Wang, Y.; Douiri, A.; Wolfe, C.D.; Bray, B. A

systematic review of machine learning models for predicting outcomes of stroke with structured data. PLoS ONE 2020, 15,

e0234722.

[16] Sirsat, M.S.; Fermé, E.; Câmara, J. Machine learning for brain stroke: A review. J. Stroke Cerebrovasc. Dis. 2020, 29,

105162.

[17] Arslan, A.K.; Colak, C.; Sarihan, M.E. Different medical data mining approaches-based prediction of ischemic stroke.

Comput. Methods Programs Biomed. 2016, 130, 87–92.

[18] Islam, M.S., Hussain, I., Rahman, M.M., Park, S.J. and Hossain, M.A., 2022. Explainable Artificial Intelligence

Model for Stroke Prediction Using EEG Signal. Sensors,22(24), p.9859.

[19] Dritsas, E. and Trigka, M., 2022. Stroke risk prediction with machine learning techniques. Sensors, 22(13), p.4670

[20] Kokkotis, C., Giarmatzis, G., Giannakou, E., Moustakidis, S., Tsatalas, T., Tsiptsios, D., Vadikolias, K. and Aggelousis, N.,

2022. An Explainable Machine Learning Pipeline for Stroke Prediction on Imbalanced Data. Diagnostics, 12(10), p.2392