STROKES NO MORE: INNOVATIVE EARLY PREDICTION USING MACHINE LEARNING

[1] Domakonda Neha, [2] Dr. V. Uma Rani, [3]Dr. Sunitha Vanamala

[1] Student, Department of Information Technology, University College of Engineering Science and Technology, JNTU Hyderabad.

[2] Professor Of CSE, Department of Information Technology, University College of Engineering Science and Technology, JNTU Hyderabad.

[3] Lecturer, Department of Computer Science, TSWRDCW, Warangal East, Warangal, Telangana, India

Abstract: Stroke poses a significant worldwide threat with severe health and economic implications, resulting from disruptions in blood flow to the brain and causing neurological impairment. As the aging population increases, the number of people at risk for stroke grows, emphasizing the urgent need for effective prediction systems. This project is addressing the challenge of stroke by developing automated prediction algorithms. These algorithms aim to enable early intervention, potentially saving lives by predicting strokes accurately. The precision and effectiveness of such systems become increasingly crucial in managing the rising population at risk. The project involves a comprehensive examination, comparing the effectiveness of a proposed machine learning technique with six well-known classifiers. Metrics related to both generalization capability and prediction accuracy were scrutinized to evaluate the performance of the developed algorithm in stroke prediction. To provide transparency into the black-box nature of machine learning models, the study employs explainable techniques, specifically SHAP (Shapley Additive Explanations). This method is well-established in the medical industry, offering insights into model decision-making processes. The experimental results indicate that more intricate models outperformed simpler ones, higher accuracy. The proposed framework, incorporating both global explainable methodology, aims to standardize complex models. This standardization can enhance stroke care and treatment by providing valuable insights into the decision-making process of the algorithms. It includes ensemble methods such as Categorical Boosting and Stacking Classifier were applied, leveraging the combined predictions of multiple individual models to enhance overall prediction accuracy. Notably, the Stacking Classifier demonstrated exceptional performance, achieving an impressive 99% accuracy.

Index terms - *Stroke prediction, data leakage, explainable machine learning.*

1. INTRODUCTION

The incidence of stroke has been increasing globally, and it is now considered one of the leading causes of death and disability. Early intervention is crucial in preventing long-term disability and mortality associated with stroke. Traditional methods of predicting stroke risk, however, are often time consuming and prone to errors. Recently, machine learning algorithms have shown great promise in accurately predicting stroke risk based on various clinical risk factors. By leveraging these algorithms, clinicians can identify high-risk patients and intervene early, potentially

reducing the number of stroke-related complications and improving patient outcomes. Additionally, there is a growing need for transparency and explainability in machine learning models in healthcare. The use of an interpretable machine learning model can provide clinicians with valuable insights into the factors that contribute to a patient's stroke risk, thereby aiding in treatment decisions. The World Stroke Organisation estimates that 13 million people worldwide experience a stroke each year, leading to 5.5 million fatalities [1]. Stroke affects all aspects of a patient's life, including their family, social environment, and work, and is one of the top causes of mortality and disability in the world [1], [2]. A common misconception is that certain groups of people, such as the elderly or those with underlying illnesses, are the only ones who are affected by stroke. In reality, anybody can be impacted, regardless of age, gender, or physical health [1], [2]. A stroke is a rapid, serious disruption in blood flow to the brain that deprives brain cells of oxygen. It comes in ischemic and hemorrhagic varieties. Moderate to severe strokes can cause permanent or temporary damage, depending on their severity. Hemorrhagic strokes are uncommon; however, they are brought on by the rupture of a blood vessel in the brain. The most common type of stroke happens when an artery is blocked or narrows, preventing blood flow to the brain [3], [4]. Age over 55, prior stroke or TIA, arrhythmia, high blood pressure, carotid stenosis from atherosclerosis, smoking, high blood cholesterol, diabetes, obesity, inactivity, estrogen therapy, blood clotting disorders, cocaine or amphetamine use, and heart issues like infarction and cardiac arrest are all risk factors for stroke [5], [6], [7]. Strokes can occur suddenly, and their symptoms might vary and be unanticipated. The main symptoms of a stroke include paralysis on one side of the body, numbness in the face, arms, or legs, difficulty speaking or walking, dizziness, blurred vision, headache, vomiting, drooping mouth, and, in severe cases, loss of consciousness and coma. These sensations may come on suddenly or gradually, and in certain rare cases, they may cause you to become aware [8], [9], [10].

2. LITERATURE SURVEY

[7] The paper, " Stroke risk factors, genetics, and prevention," addresses the Stroke is a heterogeneous syndrome, and determining risk factors and treatment depends on the specific pathogenesis of stroke. Risk factors for stroke can be categorized as modifiable and nonmodifiable. Age, sex, and race/ethnicity are nonmodifiable risk factors for both ischemic and hemorrhagic stroke, while hypertension, smoking, diet, and physical inactivity are among some of the more commonly reported modifiable risk factors. More recently described risk factors and triggers of stroke include inflammatory disorders, infection, pollution, and cardiac atrial disorders independent of atrial fibrillation. Single-gene disorders may cause rare, hereditary disorders for which stroke is a primary manifestation. Recent research also suggests that common and rare genetic polymorphisms can influence risk of more common causes of stroke, due to both other risk factors and specific stroke mechanisms, such as atrial fibrillation. Genetic factors, particularly those with environmental interactions, may be more modifiable than previously recognized. Stroke prevention has generally focused on modifiable risk factors. Lifestyle and behavioral modification, such as dietary changes or smoking cessation, not only reduces stroke risk, but also reduces the risk of other cardiovascular diseases. Other prevention strategies include identifying and treating medical conditions, such as hypertension and diabetes, that increase stroke risk. Recent research into risk factors and genetics of stroke has not only identified those at risk for stroke but also identified ways to target at-risk populations for stroke prevention.

[12] The paper, " Natural language processing and machine learning for identifying incident stroke from electronic health records: Algorithm development and validation," addresses the Stroke is an important clinical outcome in cardiovascular research. However, the ascertainment of incident stroke is typically accomplished via time-consuming manual chart abstraction. Current phenotyping efforts using electronic health records for stroke focus on case ascertainment rather than incident disease, which requires knowledge of the temporal sequence of events. The aim of this study was to develop a machine learning–based phenotyping algorithm for incident stroke ascertainment based on diagnosis codes, procedure codes, and clinical concepts extracted from clinical notes using natural language processing. The algorithm was trained and validated using an existing epidemiology cohort consisting of 4914 patients with atrial fibrillation (AF) with manually curated incident stroke events. Various combinations of feature sets and machine learning classifiers were compared. Using a heuristic rule based on the composition of concepts and codes, we further detected the stroke subtype (ischemic stroke/transient ischemic attack or hemorrhagic stroke) of each identified stroke. The algorithm was further validated using a cohort (n=150) stratified sampled from a population in Olmsted County, Minnesota (N=74,314).

[13] The paper titled " Multi-frequency symmetry difference electrical impedance tomography with machine learning for human stroke diagnosis" addresses the Multi-frequency symmetry difference electrical impedance tomography (MFSD-EIT) can robustly detect and identify unilateral perturbations in symmetric scenes. Here, an investigation is performed to assess if the algorithm can be successfully applied to identify the aetiology of stroke with the aid of machine learning. Anatomically realistic four-layer finite element method models of the head based on stroke patient images are developed and used to generate EIT data over a 5 Hz–100 Hz frequency range with and without bleed and clot lesions present. Reconstruction generates conductivity maps of each head at each frequency. Application of a quantitative metric assessing changes in symmetry across the sagittal plane of the reconstructed image and over the frequency range allows lesion detection and identification. The algorithm is applied to both simulated and human (n = 34 subjects) data. A classification algorithm is applied to the metric value in order to differentiate between normal, haemorrhage and clot values. An average accuracy of 85% is achieved when MFSD-EIT with support vector machines (SVM) classification is used to identify and differentiate bleed from clot in human data, with 77% accuracy when differentiating normal from stroke in human data. Applying a classification algorithm to metrics derived from MFSD-EIT images is a novel and promising technique for detection and identification of perturbations in static scenes. The MFSD-EIT algorithm used with machine learning gives promising results of lesion detection and identification in challenging conditions like stroke. The results imply feasible translation to human patients.

[14] The paper titled " Artificial intelligence for decision support in acute stroke—Current roles and potential" addresses The identification and treatment of patients with stroke is becoming increasingly complex as more treatment options become available and new relationships between disease features and treatment response are continually discovered. Consequently, clinicians must constantly learn new skills (such as clinical evaluations or image interpretation), stay up to date with the literature and incorporate advances into everyday practice. The use of artificial intelligence (AI) to support clinical decision making could reduce inter-rater variation in routine clinical practice and facilitate the extraction of vital information that could improve identification of patients with stroke,

prediction of treatment responses and patient outcomes. Such support systems would be ideal for centres that deal with few patients with stroke or for regional hubs, and could assist informed discussions with the patients and their families. Moreover, the use of AI for image processing and interpretation in stroke could provide any clinician with an imaging assessment equivalent to that of an expert. However, any AI-based decision support system should allow for expert clinician interaction to enable identification of errors (for example, in automated image processing). In this Review, we discuss the increasing importance of imaging in stroke management before exploring the potential and pitfalls of AI-assisted treatment decision support in acute stroke.

[15] The paper " A systematic review of machine learning models for predicting outcomes of stroke with structured data" addresses the Machine learning (ML) has attracted much attention with the hope that it could make use of large, routinely collected datasets and deliver accurate personalised prognosis. The aim of this systematic review is to identify and critically appraise the reporting and developing of ML models for predicting outcomes after stroke. Methods We searched PubMed and Web of Science from 1990 to March 2019, using previously published search filters for stroke, ML, and prediction models. We focused on structured clinical data, excluding image and text analysis. This review was registered with All studies evaluated discrimination with thirteen using area under the ROC curve whilst calibration was assessed in three. Two studies performed external validation. None described the final model sufficiently well to reproduce it. The use of ML for predicting stroke outcomes is increasing. However, few met basic reporting standards for clinical prediction tools and none made their models available in a way which could be used or evaluated. Major improvements in ML study conduct and reporting are needed before it can meaningfully be considered for practice.

3. METHODOLOGY

i) Proposed Work:

The proposed system involves evaluating stroke prediction models with a new technique and comparing it to six classifiers. Using SHAP ,the project gains insights into the decision-making process. Preprocessing and balancing the dataset with SMOTE improve accuracy. Uniquely, the project develops, explores various machine learning algorithms for early stroke prediction. The other algorithm of the study is, ensemble methods such as Categorical Boosting and Stacking Classifier were applied, leveraging the combined predictions of multiple individual models to enhance overall prediction accuracy. Notably, the Stacking Classifier demonstrated exceptional performance, achieving an impressive 99% accuracy. As a practical application of this advancement, a front end utilizing the Flask framework is developed to facilitate user testing. Additionally, integrating user authentication ensures secure access to the system, fostering a reliable and user-friendly interface for leveraging the automated stroke prediction algorithms.

ii) System Architecture:

Due to its efficiency in analyzing massive volumes of medical data, including photos of skin lesions, machine learning is being utilized more and more in medical diagnostics, including the categorization of skin cancer. The main objectives of employing machine learning models in the context of stroke prediction are to increase diagnostic precision and classification efficiency. Various machine learning models are often used to create an automated

stroke prediction system, which is then assessed using metrics like accuracy, recall, and F1 score to find the best model for the job. This study's method for categorizing stroke predictions automatically entails making a ''Yes'' or ''No'' prediction. A five-step approach is used to create the model, as illustrated in Figure 1: Getting a dataset of electronic health records is step one. Steps two and three involve pre-processing the dataset by rescaling and normalizing the data, step four involves extracting features, step five involves building a classifier algorithm using the extracted feature vectors, and step six involves using the SHAP and LIME methods to shed light on the model's decision-making process. This improved strategy attempts to increase the precision of stroke prediction and assist medical practitioners in making more knowledgeable treatment decisions.

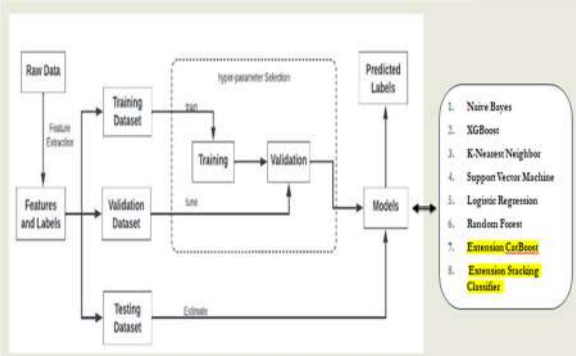

Fig 1 Proposed architecture

iii) Dataset collection:

The "Stroke Dataset" containg health-related information, including gender, age, hypertension, heart disease, marital status, occupation, residence type, average glucose level, BMI, smoking status, and stroke occurrence. Each entry is uniquely identified by an 'id.' The dataset provides valuable insights into the factors associated with stroke, with binary indicators for hypertension, heart disease, and marital and residence status. This dataset is loaded and explored. This includes examining the structure of the dataset, checking for missing values, and gaining insights into the distribution and characteristics of the features.

	id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
0	9046	Male	67.0	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
1	51676	Female	61.0	0	0	Yes	Self-employed	Rural	202.21	0.0	never smoked	1
2	31112	Male	80.0	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
3	60182	Female	49.0	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
4	1665	Female	79.0	1	0	Yes	Self-employed	Rural	174.12	24.0	never smoked	1
...	...	...	...	...	...	...	...	...	...	...	...	...
5105	18234	Female	80.0	1	0	Yes	Private	Urban	83.75	0.0	never smoked	0
5106	44873	Female	81.0	0	0	Yes	Self-employed	Urban	125.20	40.0	never smoked	0
5107	19723	Female	35.0	0	0	Yes	Self-employed	Rural	82.99	30.6	never smoked	0
5108	37544	Male	51.0	0	0	Yes	Private	Rural	166.29	25.6	formerly smoked	0
5109	44679	Female	44.0	0	0	Yes	Govt_job	Urban	85.28	26.2	Unknown	0

Fig 2 Stroke dataset

iv) Data Processing:

Data processing involves transforming raw data into valuable information for businesses. Generally, data scientists process data, which includes collecting, organizing, cleaning, verifying, analyzing, and converting it into readable

formats such as graphs or documents. Data processing can be done using three methods i.e., manual, mechanical, and electronic. The aim is to increase the value of information and facilitate decision-making. This enables businesses to improve their operations and make timely strategic decisions. Automated data processing solutions, such as computer software programming, play a significant role in this. It can help turn large amounts of data, including big data, into meaningful insights for quality management and decision-making.

v) Feature selection:

Feature selection is the process of isolating the most consistent, non-redundant, and relevant features to use in model construction. Methodically reducing the size of datasets is important as the size and variety of datasets continue to grow. The main goal of feature selection is to improve the performance of a predictive model and reduce the computational cost of modeling.

Feature selection, one of the main components of feature engineering, is the process of selecting the most important features to input in machine learning algorithms. Feature selection techniques are employed to reduce the number of input variables by eliminating redundant or irrelevant features and narrowing down the set of features to those most relevant to the machine learning model. The main benefits of performing feature selection in advance, rather than letting the machine learning model figure out which features are most important.

vi) Algorithms:

XGBoost is an advanced machine learning algorithm categorized under gradient boosting frameworks. It excels in both regression and classification tasks, utilizing an ensemble approach to sequentially build decision trees that correct errors. Its "extreme" capabilities lie in its efficiency, scalability, and regularization techniques. XGBoost is utilized in the project for its superior predictive performance, efficiently handling complex relationships within clinical risk factors associated with strokes. Its ensemble approach ensures accurate predictions, aligning with the project's goal of developing a precise tool for early identification and intervention in high-risk stroke patients.

```
#now train XGBoost algorithm
xg_cls =  XGBClassifier(n_estimators=10)#define XGBOOST object
xg_cls.fit(X_train, y_train)#train XGBOost on training data
predict = xg_cls.predict(X_test)#perform prediction on test data
calculateMetrics("XGBoost", predict, y_test)#calculate accuracy and other metrics

xgb_acc = accuracy_score(predict, y_test)
xgb_prec = precision_score(predict, y_test,average='macro')
xgb_rec = recall_score(predict, y_test,average='macro')
xgb_f1 = f1_score(predict, y_test,average='macro')

storeResults('XGBoost',xgb_acc,xgb_prec,xgb_rec,xgb_f1)
```

Fig 3 XGBoost

Naive Bayes is a probabilistic machine learning algorithm based on Bayes' theorem, assuming independence between features. Naive Bayes is chosen for its simplicity and efficiency in handling high-dimensional, potentially correlated data. In stroke prediction, where diverse clinical factors contribute, Naive Bayes provides a computationally efficient solution, aligning with the project's objective of accurate risk prediction. Its ease of implementation and interpretability make it a practical choice for healthcare applications.

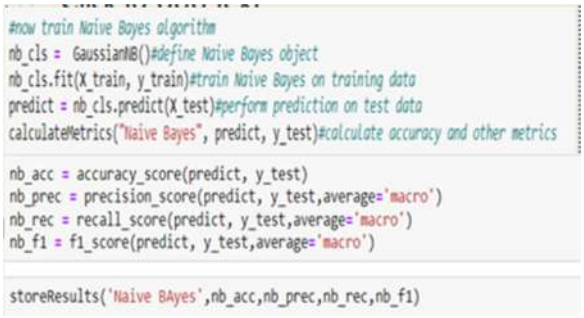

Fig 4 Naïve bayes

KNN is a versatile machine learning algorithm for classification and regression, relying on proximity principles. In stroke prediction, where clinical risk factors' relationships are complex, KNN's simplicity allows pattern identification based on data point proximity. Its adaptability to varying data distributions and handling of non-linear relationships align with the project's goal of accurately predicting stroke risk. Particularly beneficial when decision boundaries are unclear, KNN suits scenarios with non-linear or intricate data structures.

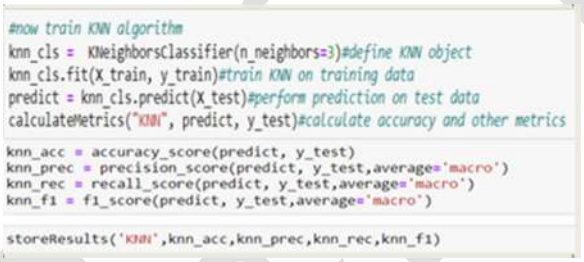

Fig 5 KNN

SVM is a powerful algorithm for classification and regression tasks, excelling in high-dimensional spaces. In the project, SVM is chosen for its effectiveness in handling complex relationships within clinical risk factors associated with stroke. By identifying optimal hyperplanes, SVM enhances precision in stroke risk prediction, particularly suitable for intricate patterns in the data. Its adaptability to high-dimensional and non-linear data aligns with the project's goal of developing an accurate predictive model.

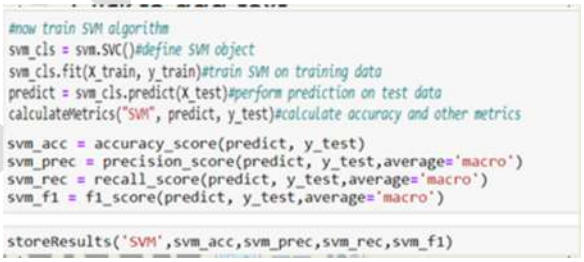

Fig 6 SVM

Logistic Regression is a statistical method for binary classification, modeling the probability of an instance belonging to a specific class using the logistic function. Chosen for simplicity and effectiveness, Logistic Regression is employed for binary classification in stroke prediction. Its straightforward approach aligns with the project's goal of developing a reliable and interpretable model to classify stroke risk based on diverse clinical features.

```
#now train LogisticRegression algorithm
lr_cls = LogisticRegression()#define regression object
lr_cls.fit(X_train, y_train)#train regression on training data
predict = lr_cls.predict(X_test)#perform prediction on test data
calculateMetrics("Logistic Regression", predict, y_test)#calculate accuracy and other metrics

lr_acc = accuracy_score(predict, y_test)
lr_prec = precision_score(predict, y_test,average='macro')
lr_rec = recall_score(predict, y_test,average='macro')
lr_f1 = f1_score(predict, y_test,average='macro')

storeResults('Logistic Regression',lr_acc,lr_prec,lr_rec,lr_f1)
```

Fig 7 Logistic regression

Random Forest is an ensemble learning algorithm that aggregates predictions from multiple decision trees for classification or regression tasks. Chosen for its ability to handle complex relationships in clinical risk factors, Random Forest enhances accuracy by combining predictions from numerous trees. Its effectiveness in managing high-dimensional data and preventing overfitting aligns with the project's goal of developing a highly accurate and generalizable predictive model for stroke risk.

```
#train random forest algorithm on training dataset and test its prediction capability on test data
#now train Random Forest algorithm
rf_cls = RandomForestClassifier()
rf_cls.fit(X_train, y_train)
predict = rf_cls.predict(X_test)
calculateMetrics("Random Forest", predict, y_test)

lr_acc = accuracy_score(predict, y_test)
lr_prec = precision_score(predict, y_test,average='macro')
lr_rec = recall_score(predict, y_test,average='macro')
lr_f1 = f1_score(predict, y_test,average='macro')

storeResults('Logistic Regression',lr_acc,lr_prec,lr_rec,lr_f1)
```

Fig 8 Random forest

A **Stacking Classifier** is an ensemble technique combining multiple classifiers to enhance predictive performance by using a meta-classifier. Employed to leverage diverse strengths of classifiers. The Stacking Classifier aims to create a powerful and accurate stroke risk prediction model by combining outputs from these models, addressing individual algorithm weaknesses for comprehensive patient risk identification.

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import DecisionTreeClassifier
from lightgbm import LGBMClassifier
from sklearn.ensemble import StackingClassifier

estimators = [('rf', RandomForestClassifier(n_estimators=10)),('dt', DecisionTreeClassifier())]

clf = StackingClassifier(estimators=estimators, final_estimator=LGBMClassifier())

# fit the model
clf.fit(X_train, y_train)

y_pred = clf.predict(X_test)

stac_acc_a = accuracy_score(y_test,y_pred)
stac_prec_a = precision_score(y_test,y_pred)
stac_rec_a = recall_score(y_test,y_pred)
stac_f1_a = f1_score(y_test,y_pred)

calculateMetrics("Stacking", predict, y_test)#calculate accuracy and other metrics
```

Fig 9 Stacking classifier

CatBoost is a powerful gradient boosting algorithm designed for decision trees, known for efficient handling of categorical features without extensive preprocessing. Chosen for its excellence with categorical features, CatBoost streamlines modeling, minimizing preprocessing efforts. Its efficiency and robustness contribute to accurate stroke risk predictions, capturing complex relationships within clinical risk factors. CatBoost enhances precision and generalization capability in stroke prediction.

```
#now train extension CATBOOST algorithm as extension which is more advanced then other ML algorithm
cb_cls = cb.CatBoostClassifier(iterations=300, learning_rate=0.1)
cb_cls.fit(X_train, y_train)#train CatBoost on training data
predict = cb_cls.predict(X_test)#perform prediction on test data
calculateMetrics("Extension CatBoost", predict, y_test)#calculate accuracy and other metrics

cat_acc = accuracy_score(predict, y_test)
cat_prec = precision_score(predict, y_test,average='macro')
cat_rec = recall_score(predict, y_test,average='macro')
cat_f1 = f1_score(predict, y_test,average='macro')

storeResults('CatBoost',cat_acc,cat_prec,cat_rec,cat_f1)
```

Fig 10 Catboost

4. EXPERIMENTAL RESULTS

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

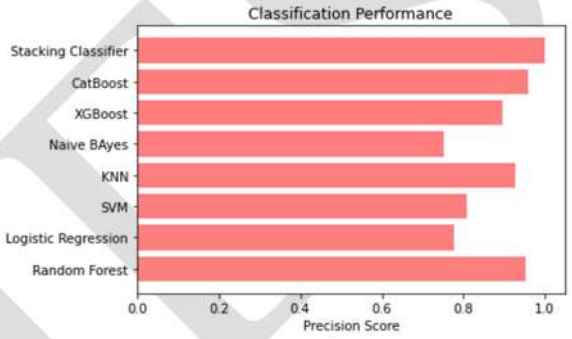

Fig 11 Precision comparison graph

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{TP}{TP + FN}$$

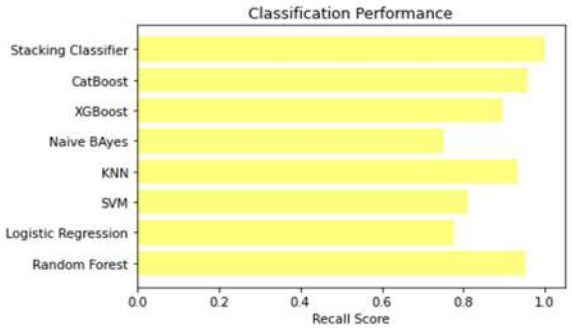

Fig 12 Recall comparison graph

Accuracy: Accuracy is the proportion of correct predictions in a classification task, measuring the overall correctness of a model's predictions.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

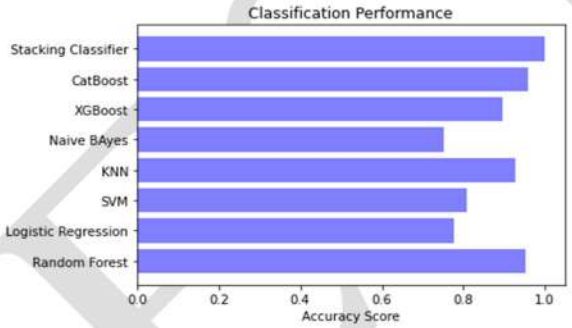

Fig 13 Accuracy graph

F1 Score: The F1 Score is the harmonic mean of precision and recall, offering a balanced measure that considers both false positives and false negatives, making it suitable for imbalanced datasets.

$$\text{F1 Score} = 2 * \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} * 100$$

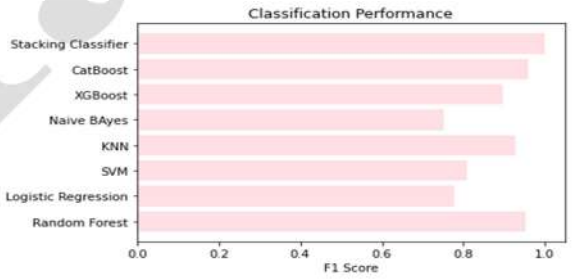

Fig 14 F1Score

ML Model	Accuracy	Precision	F1_score	Recall
Random Forest	0.952	0.952	0.952	0.952
Logistic Regression	0.777	0.777	0.777	0.778
SVM	0.809	0.808	0.808	0.812
KNN	0.927	0.926	0.927	0.932
Naive Bayes	0.752	0.751	0.751	0.752
XGBoost	0.895	0.895	0.895	0.897
Extension CatBoost	0.960	0.960	0.960	0.960
Extension Stacking Classifier	0.999	0.999	0.999	0.999

Fig 15 Performance Evaluation

Fig 16 Home page

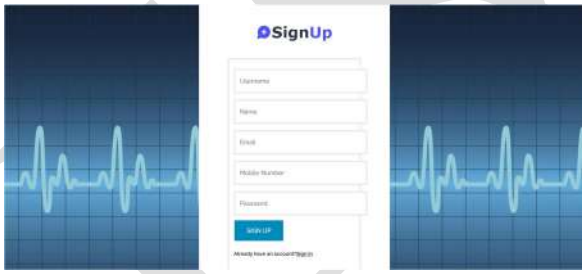

Fig 17 Signup page

Fig 18 Login page

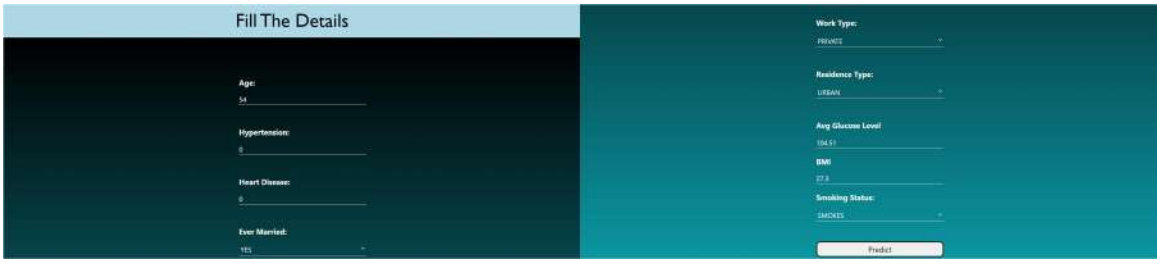

Fig 19 User input

Fig 20 Predict result for given input

5. CONCLUSION

Through the project, significant strides have been made in enhancing stroke prediction accuracy, leveraging state-of-the-art machine learning algorithms for early identification of high-risk individuals. The project successfully addressed class imbalance issues within the dataset, ensuring a more robust and balanced representation of stroke and non-stroke cases for improved model performance. The extension algorithm, Stacking Classifier, excelled with a 99% accuracy in stroke prediction. Successfully integrated into a user-friendly front end, it processed feature values accurately, showcasing both robust performance and practical applicability for real-world healthcare implementation. By standardizing complex models using global and local explainable methodologies, the project contributes to a more standardized and effective approach in stroke care, promoting better treatment strategies across diverse patient scenarios. The creation of trustworthy and transparent AI systems, offering concise and interpretable explanations, marks a significant contribution to the field of Explainable Artificial Intelligence (XAI) in the medical domain, ensuring accountable and reliable predictive models for stroke risk assessment.

6. FUTURE SCOPE

Future research could focus on exploring a wider array of machine learning algorithms and techniques to enhance the accuracy and generalization capabilities of stroke prediction models. The development of the web application for early stroke intervention offers potential for growth. Future endeavors could include the addition of advanced features and functionalities to elevate the impact on stroke care and treatment. The integration of global and local explainable methodologies, especially SHAP, presents opportunities for further study. Investigating these methods could provide deeper insights into the decision-making processes of machine learning models used in stroke prediction. A prospective direction involves creating a comprehensive end-to-end smart stroke prediction system, extending its capabilities to mobile applications for both Android and iOS platforms. This expansion could increase accessibility and usability. Exploring demographic factors like age and gender in more detail is a promising avenue for future research. Understanding their nuanced impact on stroke occurrence could pave the way for the development of more personalized and targeted prediction models.

REFERENCES

[1] Learn About Stroke. Accessed: May 25, 2022. [Online]. Available: https://www.world-stroke.org/world-stroke-day-campaign/why-strokematters/learn-about-stroke

[2] T. Elloker and A. J. Rhoda, ‘‘The relationship between social support and participation in stroke: A systematic review,’’ Afr. J. Disability, vol. 7, pp. 1–9, Oct. 2018.

[3] M. Katan and A. Luft, ‘‘Global burden of stroke,’’ Seminar Neurol., vol. 38, no. 2, pp. 208–211, Apr. 2018.

[4] A. Bustamante, A. Penalba, C. Orset, L. Azurmendi, V. Llombart, A. Simats, E. Pecharroman, O. Ventura, M. Ribó, D. Vivien, J. C. Sanchez, and J. Montaner, ‘‘Blood biomarkers to differentiate ischemic and hemorrhagic strokes,’’ Neurology, vol. 96, no. 15, pp. e1928–e1939, Apr. 2021.

[5] X. Xia, W. Yue, B. Chao, M. Li, L. Cao, L. Wang, Y. Shen, and X. Li, ‘‘Prevalence and risk factors of stroke in the elderly in northern China: Data from the national stroke screening survey,’’ J. Neurol., vol. 266, no. 6, pp. 1449–1458, Jun. 2019.

[6] A. Alloubani, A. Saleh, and I. Abdelhafiz, ‘‘Hypertension and diabetes mellitus as a predictive risk factors for stroke,’’ Diabetes Metabolic Syndrome, Clin. Res. Rev., vol. 12, no. 4, pp. 577–584, Jul. 2018.

[7] A. K. Boehme, C. Esenwa, and M. S. V. Elkind, ‘‘Stroke risk factors, genetics, and prevention,’’ Circ. Res., vol. 120, no. 3, pp. 472–495, Feb. 2018.

[8] I. Mosley, M. Nicol, G. Donnan, I. Patrick, and H. Dewey, ‘‘Stroke symptoms and the decision to call for an ambulance,’’ Stroke, vol. 38, no. 2, pp. 361–366, Feb. 2007.

[9] J. Lecouturier, M. J. Murtagh, R. G. Thomson, G. A. Ford, M. White, M. Eccles, and H. Rodgers, ‘‘Response to symptoms of stroke in the UK: A systematic review,’’ BMC Health Services Res., vol. 10, no. 1, pp. 1–9, Dec. 2010.

[10] L. Gibson and W. Whiteley, ‘‘The differential diagnosis of suspected stroke: A systematic review,’’ J. Roy. College Physicians Edinburgh, vol. 43, no. 2, pp. 114–118, Jun. 2013.

[11] N. M. Murray, M. Unberath, G. D. Hager, and F. K. Hui, ‘‘Artificial intelligence to diagnose ischemic stroke and identify large vessel occlusions: A systematic review,’’ J. NeuroInterventional Surgery, vol. 12, no. 2, pp. 156–164, Feb. 2020.

[12] Y. Zhao, S. Fu, S. J. Bielinski, P. A. Decker, A. M. Chamberlain, V. L. Roger, H. Liu, and N. B. Larson, ‘‘Natural language processing and machine learning for identifying incident stroke from electronic health records: Algorithm development and validation,’’ J. Med. Internet Res., vol. 23, no. 3, Mar. 2021, Art. no. e22951.

[13] B. McDermott, A. Elahi, A. Santorelli, M. O’Halloran, J. Avery, and E. Porter, ‘‘Multi-frequency symmetry difference electrical impedance tomography with machine learning for human stroke diagnosis,’’ Physiological Meas., vol. 41, no. 7, Aug. 2020, Art. no. 075010.

[14] A. Bivard, L. Churilov, and M. Parsons, ‘‘Artificial intelligence for decision support in acute stroke—Current roles and potential,’’ Nature Rev. Neurol., vol. 16, no. 10, pp. 575–585, Oct. 2020.

[15] W. Wang, M. Kiik, N. Peek, V. Curcin, I. J. Marshall, A. G. Rudd, Y. Wang, A. Douiri, C. D. Wolfe, and B. Bray, ‘‘A systematic review of machine learning models for predicting outcomes of stroke with structured data,’’ PLoS ONE, vol. 15, no. 6, Jun. 2020, Art. no. e0234722.