Machine Learning-Based Automated Stroke Prediction: An Explanatory and Investigative Study Using a Web Application for Prevention

[1]**Saranya K,** M. Tech (PG Student), Department of Computer Science and Engineering, Joginpally B.R.Engineering College, saranyait011@gmail.com

[2]**Dr. D. Magdalene Delighta Angeline,** Associate Professor, Department of Computer Science and Engineering, Joginpally B.R. Engineering College, magdalenedelighta@gmail.com

Abstract: Stroke represents a massive international hazard, with profound fitness and financial consequences, Cerebral blood flow is obstructed, leading to neurological impairments. The green prediction systems stress the importance of stroke in persons who are at risk of having one, especially in older people who are getting older. This study addresses the challenges of ironing by enhancing future automated algorithms. These algorithms need to make it easier to intervene early, and they will definitely save lives by predicting strokes. There are a lot of people that want to prevent population growth because this technology is so accurate and effective. The inspection entails a comprehensive examination, resulting in a warning machine learning technique with six distinct classes. The overall effectiveness of the established criteria in stroke prediction changed into assessed with the aid of using inspecting metrics touching on each generalization functionality and prediction accuracy. The observe makes use of explainable methodologies, appreciably SHAP (Shapley Additive Explanations), to explain the opaque nature of ML fashions. This approach is well-hooked up withinside the scientific field, To provide insight into the decision-making processes for the version. Experimental data suggest that increasingly intricate outfits surpass less complex individuals in accuracy. The warning structure, which includes a worldwide explanatory method, intends to make detailed fashion the same across the board. This standardization can enhance stroke care and treatment through the algorithm's decision-making processes. The dress strategy, which includes graded boosting and stacking classifications, is used to make the overall prediction more accurate. This is done by using pool conditions of different fashions. The stacking classifier did a great job overall, with an accuracy of 99%.

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

1. INTRODUCTION

Stroke is now one of the most important reasons of demise and incapacity across the world, and its variety has been rising. Early remedy could be very essential for fending off long-time period incapacity and demise after a stroke. But conventional approaches of figuring out the risk of having a stroke take a long time and often make mistakes. Recently, ML algorithms have shown a lot of promise in being able to correctly predict the risk of stroke based on different clinical risk factors. Using these algorithms, doctors can find patients who are at high risk and help them right away, which could lower the number of complications related to strokes and improve patient results. Also, there is a growing need for ML models in healthcare to be clear and easy to understand. Using a ML model that can be interpreted can help doctors figure out what makes a patient more likely to have a stroke, which can help them decide how to treat them. According to stroke organizations around the world, 13 million people suffer strokes each year, and 5.5 million people die from them [1]. Strokes are one of the major reasons for death and disability worldwide [1, 2]. It changes every part of a person's life, whether it's family, friends, work, or anything else. There is a common misconception that strokes only occur in certain types of people, such as elderly people or those who already have health problems. In fact, it can affect everyone regardless of age, gender, or health [1, 2]. When blood flow to the brain is suddenly and seriously blocked, brain cells lose oxygen. This is called a stroke. It comes in two types: cerebral and hemorrhagic. Depending on how bad they are, moderate to serious strokes can damage you permanently or temporarily. Not many people have hemorrhagic strokes, but they happen when Blood vessels break in the brain. Most strokes occur when the artery stops or narrows and blood flow is blocked by the brain [3,4]. A person is over 55 years old and suffers a stroke before a stroke or TIA. There are issues like arrhythmia, smoke, high blood cholesterol, diabetes, overweight, activity, estrogen, estrogen, blood clotting problems, cocaine or amphetamine problems, or heart placement. [5, 6, 7]. Strokes can occur soon, and the signs can change and become strange times. Some of the main symptoms of a stroke are paralysis of the body, numbness of the face, problems with speech or walking, dizziness, stains, headaches, vomiting, an open mouth, fainting, and going into a coma. These might happen quickly or over time, and in extremely rare circumstances, they can get people's attention [8, 9, 10].

2. LITERATURE SURVEY

This article, "Stroke chance elements, genetics, and prevention," talks approximately the As a symptom, stroke may be very one of a kind from one character to any other. The purpose of a stroke determines chance elements and treatment. There are forms of chance elements for stroke: people who may be modified and people that can`t. Age, sex, and race or ethnicity are chance elements for each ischemic and hemorrhagic stroke that cannot be modified. On the alternative hand, excessive blood strain, smoking, diet, and absence of bodily workout are a number of the maximum not unusualplace chance elements that may be modified. Inflammatory problems, infections, pollution, and coronary heart atrial problems that aren't as a result of atrial traumatic inflammation were named as new chance elements and reasons of stroke. Stroke is regularly the primary signal of uncommon genetic sicknesses which are as a result of a unmarried gene. New take a look at additionally suggests that each not unusualplace and uncommon genetic polymorphisms can have an effect on the chance of extra not unusualplace reasons of stroke. This is due to the fact they are able to have an effect on each different chance elements and particular stroke mechanisms, like atrial traumatic inflammation. Genetic elements, specifically people who hook up with the environment, can be simpler to extrade than become idea before. Stroke prevention has in general been targeted on chance elements that may be modified. Making modifications for your life-style and habits, like quitting smoking or looking what you eat, can decrease your chance of stroke and different coronary heart sicknesses as well. Finding and treating clinical issues like excessive blood strain and diabetes that enhance the chance of stroke is any other manner to prevent them. New study into the genetics and risk factors of stroke has not only found people who are likely to have a stroke, but also ways to prevent strokes in those groups.

[12] Stroke is a significant clinical outcome in cardiac research, and this work aims to find phenomena related to brain function through NLP and ML. It discusses about electronic health records, algorithm creation, and verification. Finding out about an incident stroke, on the other hand, is usually done by manually abstracting charts, which takes a lot of time. The way electronic health records are used now for phenotyping strokes is mainly for finding cases, not for finding incident diseases, which needs knowing the order of events in time. This study was to create a phenotypic algorithm to find the type of stroke created in diagnostic codes, procedural codes, and clinical ideas based on clinical notes using NLP and ML. An existing epidemiological group of,91

patients from Atrium Preats (AF) was used for carefully curated training and testing. Various mixtures of feature sets and machine learning models were seen side by side. Rules were used based on ideas and code structure to determine the type of "stroke (ischemic stroke, temporary ischemic attack or hemorrhagic stroke) for all. A cohort (n=150) stratified sample from a group in Olmsted County, Minnesota (N=74,314) was used to test the algorithm even more".

[13] "Multi-frequency symmetry difference electrical impedance tomography with ML for human stroke diagnosis" is the title of a paper that We'll talk about how multifrequency symmetric differential differential dance tomography (MFSD-OIT) can reliably find and invoke changes on one side in a symmetric scene. Here the investigation is conducted and the algorithm is used to determine whether the cause of the stroke can be properly found with the help of ML. Anatomically accurate and stroke-based model of heads with four-layer finite element methods can generate EIT data from patient photographs and create frequency ranges from 5 Hz to 100 Hz with and without blood and clotting lesions. The management card for each head at all frequencies is created by reconfiguration. Using quantitative measures to investigate changes in sagittal level symmetry in reconstructed images and frequency ranges, lesions are found and named. The method is used on both fake data and data from 34 real people. The metric value is put through a classification program to tell the difference between "normal, hemorrhage, and clot values. MFSD-EIT with support vector machines (SVM)" classification can tell the difference between a bleed and a "clot in human data with an average accuracy of 85%". It can also tell the difference between a normal blood flow and a stroke in

human data with a 77% accuracy. Using a classification algorithm on metrics obtained from MFSD-EIT images is a new and interesting way to find and name changes in scenes that are otherwise still. When used with ML, the MFSD-EIT method shows promise for finding and identifying lesions in difficult situations like stroke. The results show that it is possible to apply the method to real cases.

[14] A paper called AI to support decision support in acute stroke - rechasing potential conversations with rollers on this topic. There are increasing numbers of treatment decisions in stroke patients, and new connections are constantly seen between disease characteristics and treatment response. This makes it harder to diagnose and treat stroke patients. That's why doctors have to DL new things, like how to do clinical evaluations or read images, read the latest research, and use these new skills in their daily work. Using AI to help doctors make decisions could cut down on differences between raters in everyday clinical practice and make it easier to get important data that could help find stroke patients, predict how they will respond to treatment, and improve patient outcomes. These kinds of support systems would work well in centers that only see a few stroke patients or in hubs across the area. They could help patients and their families have more informed conversations. Also, using AI to process and understand stroke images could give any doctor an imaging report that is just as good as one from an expert. “Any AI-based decision support system, though, should let an expert clinician work with it so that mistakes can be found (for example, in automated image processing)”. This review talks about how imaging is becoming more and more important in managing strokes. It then looks at the pros and cons of using AI to help make treatment decisions for people who have recently had a stroke.

[15] The study "A systematic review of ML models for predicting outcomes of stroke with structured data" tackles the A lot of people are interested in ML because they think it can use big, regularly collected datasets to give accurate, personalized prognoses. The purpose of this systematic study is to find and evaluate the report and development of ML models for prediction, and to assess what happens after a stroke. The way to do that was confirmed by “PubMed and Web of Science from 1990 to March 2019 using search terms already published for Stroke, ML, and predictive models”. We only dealt with organized clinical data, not with photographs or text. This review was preserved under the ROC curve and in 13 studies investigated to examine classification, three viewed calibrations. External validation was performed in two cases. None of them provided sufficient information about the final model to make copies of it. ML is increasingly being used to predict how strokes occur. However, only a few of them followed the basic rules for reporting clinical prediction tools, and were not public enough to be used or evaluated for those models. Before seriously considering using ML, many tasks need to be applied to the way research is carried out and how they are reported.

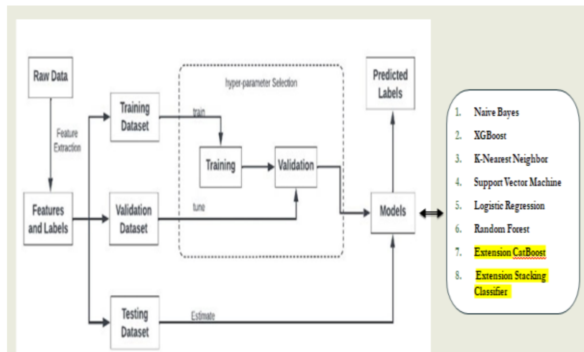
3. METHODOLOGY

i) Proposed Work:

A new method is used to test stroke prediction models and see how they compare to six classifiers in the suggested system. By using SHAP, the group learns more about how decisions are made. The accuracy is better when the information is preprocessed and balanced with SMOTE. In a unique way, the project creates and tests different ML methods for predicting strokes early on. The other algorithm used in the study was ensemble methods, which use the predictions of multiple separate models to make the general prediction more accurate. These methods include Categorical Boosting and Stacking Classifier. In particular, the stacking classifier worked surprisingly well, achieving an impressive 99 course. An example of this new technology is to create a front-end in the flask framework to simplify the test user. Additionally, by adding user authentication, you can only access the system and create a secure, user-friendly interface for using automated hub prediction methods.

ii) System Architecture:

ML is being used more and more in medical diagnostics, like putting skin cancer into groups, because it is good at processing huge amounts of medical data, like shots of skin lesions. Using ML models to identify strokes is mostly done to improve the accuracy of diagnoses and the speed with which patients are grouped. A system that automatically predicts strokes is often made using a number of different ML models. This system is then tested using accuracy, memory, and F1 score to find the best model for the job. Making a "Yes" or "No" guess is how this study automatically sorts stroke predictions into groups. There are five steps used to make the model, which can be seen in Figure 1: The first step is to get a set of electronic health information. The first three steps prepare the data records through new calm and normalization. The fourth step is to extract the features. The fifth step is to create a classification algorithm using the extracted feature vectors. The sixth step is to demonstrate how the model makes decisions using SHAP and LIME methods. This better approach tries to make it easier for physicians to make better decisions about treatment and make more accurate predictions about stroke.



"Fig 1 Proposed architecture"

iii) "Dataset collection":

The "Stroke Data Set contains health-related data such as gender, age, hypertension, heart disease, marriage status, type of work, average glucose level, BMI, smoking status", and number of strokes. Each post has a clear "ID" that inspires it. With binary indicators of hypertension, heart disease, marriage and settlement, this collection provides useful information on stroke-related things. This information has been loaded and looked into. This includes looking at the structure of the dataset, making sure there are no missing values, and learning more about how the features are distributed and what their traits are.

	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 is the process of turning data that isn't structured into useful information for the business. You usually are in charge of data scientists. This means that material is gathered, sorted, cleansed, checked, analyzed, and turned into pictures and papers that others may read. You can process data by hand, by machine, or by computer. The goal is to encourage information that is helpful and helps people make decisions. This helps organizations run more smoothly and make smarter strategic choices more quickly. Automated data processing technologies, like computer programming, make this achievable most of the time. This helps turn huge data and other kinds of data into valuable information that can be used for making decisions and checking quality.

v) "Feature selection":

Features selection is the process in which the most consistent, most convenient, and less redundant features can be used to create models. With the number and record types increasing, it is important to reduce the size in a planned manner. One of the main objectives of feature selection is to reduce the better capabilities and computing power of the predictive model.

One of the most important parts of functional engineering is the selection of characteristics, which is the process of selecting the most important features for feeding with ML algorithms. The distinctive selection method removes unnecessary or useless features and retains only the features that are most important to the ML model. This reduces the number of input variables. Instead of having your ML model do this, selecting which functions are important in advance will have the main advantage here.

vi) Algorithms:

XGBoost is a sophisticated ML method that belongs to the group of gradient boosting frameworks. It is very good at both regression and classification tasks. It does this by using an ensemble method to build decision trees in a way that fixes mistakes as it goes. Its "extreme" skills come from the way it works, how well it can scale, and how well it can regularize data. XGBoost is used in the project because it is very good at making predictions and can handle complicated links between clinical risk factors for strokes very well. Its ensemble method makes sure that predictions are correct, which is in line with the project's goal of creating a precise tool for finding high-risk stroke patients early and getting them help.

```
#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 XGBoos"t

"Naive Bayes" It is a probability ML method based on the Bayesian theorem, and assumes that the properties are not connected. Naive Bayes is chosen because it is easy to use and involves a large amount of data that can be connected. Naive Bayes is a simple method of predicting the risk of stroke based on many different clinical factors. It fits the purpose of the project to make accurate risk predictions. It's a good choice for healthcare apps because it's easy to set up and understand.

```
#now train Naive Bayes algorithm
nb_cls = GaussianNB()#define Naive Bayes object
nb_cls.fit(X_train, y_train)#train Naive Bayes on training data
predict = nb_cls.predict(X_test)#perform prediction on test data
calculateMetrics("Naive Bayes", predict, y_test)#calculate accuracy and other metrics

nb_acc = accuracy_score(predict, y_test)
nb_prec = precision_score(predict, y_test,average='macro')
nb_rec = recall_score(predict, y_test,average='macro')
nb_f1 = f1_score(predict, y_test,average='macro')

storeResults('Naive BAyes',nb_acc,nb_prec,nb_rec,nb_f1)
```

"Fig 4 Naïve bayes"

KNN is a flexible ML method that uses proximity rules to do both classification and regression. In stroke prediction, where the links between clinical risk factors are complicated, KNN's simplicity lets us find patterns by looking at how close two data points are to each other. The project's goal is to correctly predict stroke risk, and this model can handle non-linear relationships and change based on different data distributions. KNN works well in situations with complex or non-linear data structures because it helps when decision limits are not clear.

```
#now train KNN algorithm
knn_cls = KNeighborsClassifier(n_neighbors=3)#define KNN object
knn_cls.fit(X_train, y_train)#train KNN on training data
predict = knn_cls.predict(X_test)#perform prediction on test data
calculateMetrics("KNN", predict, y_test)#calculate accuracy and other metrics

knn_acc = accuracy_score(predict, y_test)
knn_prec = precision_score(predict, y_test,average='macro')
knn_rec = recall_score(predict, y_test,average='macro')
knn_f1 = f1_score(predict, y_test,average='macro')

storeResults('KNN',knn_acc,knn_prec,knn_rec,knn_f1)
```

Fig 5 KNN

SVM is a powerful method that works especially well in "high-dimensional spaces for tasks like classification and regression". SVM was picked for the project because it is good at dealing with complicated relationships between clinical risk factors for stroke. By finding the best hyperplanes, SVM improves the accuracy of predicting the risk of stroke. It works especially well when the data has complex patterns. It can handle multidimensional and nonlinear data, which fits with the project's goal of making a good prediction model.

```
#now train SVM algorithm
svm_cls = svm.SVC()#define SVM object
svm_cls.fit(X_train, y_train)#train SVM on training data
predict = svm_cls.predict(X_test)#perform prediction on test data
calculateMetrics("SVM", predict, y_test)#calculate accuracy and other metrics

svm_acc = accuracy_score(predict, y_test)
svm_prec = precision_score(predict, y_test,average='macro')
svm_rec = recall_score(predict, y_test,average='macro')
svm_f1 = f1_score(predict, y_test,average='macro')

storeResults('SVM',svm_acc,svm_prec,svm_rec,svm_f1)
```

"Fig 6 SVM"

"Logistic Regression A statistical technique for binary classification that use logistical functions to simulate the likelihood of an instance belonging to a specific category. Logistic regression" is a good choice for binary classification of stroke spread since it is simple to apply and works effectively. The goal of this research is to build a model that can accurately and easily tell how likely someone is to have a stroke based on a number of clinical factors. A simple approach to this method fits this goal.

```
#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 a type of ensemble learning method that takes predictions from several decision trees and puts them all together for tasks like classification or regression. RF improves accuracy by combining estimates from several trees. It was chosen because it can handle complex relationships in clinical risk factors. The goal of the project is to create a very accurate and usable model for predicting stroke risk, and this tool works well at handling large amounts of data and avoiding overfitting.

```
#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 method that uses a meta-classifier to combine several classifiers to improve the accuracy of predictions. Used to take advantage of the different strengths of algorithms. By combining the results of these models, the Stacking Classifier tries to make a strong and accurate stroke risk prediction model that fixes the flaws in each method for a full picture of a patient's risk.

```
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 strong gradient-boosting algorithm made for decision trees that is known for being good at working with category features without a lot of extra work. Because it is so good at category features, CatBoost speeds up modeling and reduces the amount of work that needs to be done before it can be used. Its speed and dependability help make accurate predictions about the risk of stroke by recording complex relationships between clinical risk factors. CatBoost improves the accuracy and generalizability of stroke prediction.

```
#now traon 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)
```

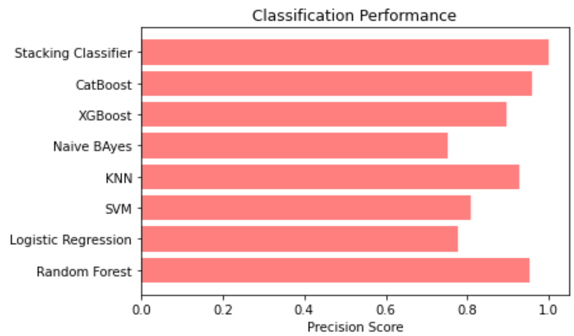
"Fig 10 Catboost"

4. EXPERIMENTAL RESULTS

"Precision": Accuracy is the number of instances or samples that were accurately categorized as positive matches compared to the total number of cases or samples. So here's how to check the accuracy:

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

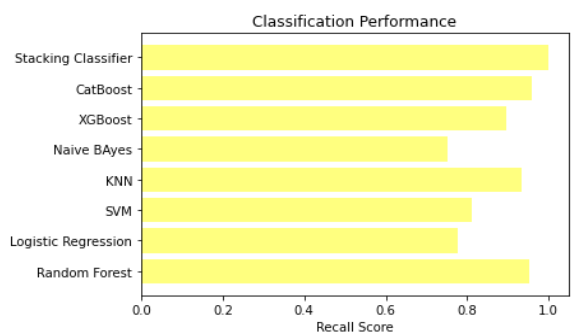
$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$



"Fig 11 Precision comparison graph"

"Recall": In ML, a callback is a metric that shows how well a model can find all the important instances of a particular class. It shows how well a particular class of models captures. This is calculated by sharing the number of positive observations correctly predicted by the actual positive total number.

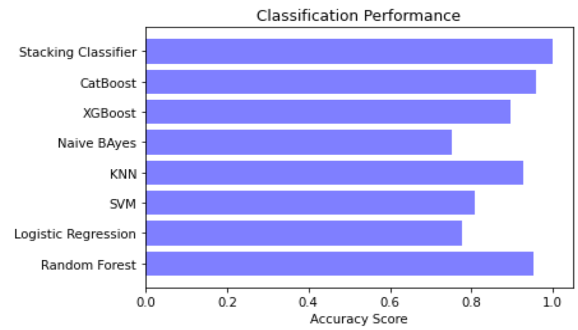
$$Recall = \frac{TP}{TP + FN}$$



"Fig 12 Recall comparison graph"

"Accuracy": Accuracy denotes the proportion of correct predictions in a classification task. It demonstrates the precision of the model's predictions.

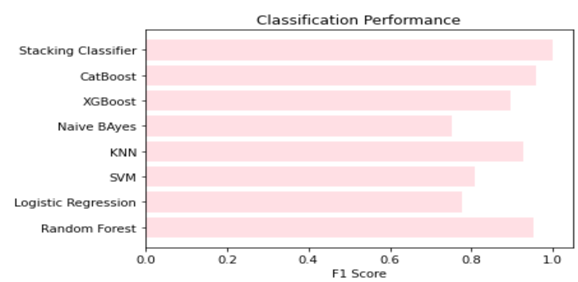
$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$



"Fig 13 Accuracy graph"

"F1 Score": F1 scores are a harmonious means of accuracy and recall. This is a fair measure that takes into account both false positives and false negatives, and can be used with unbalanced data records.

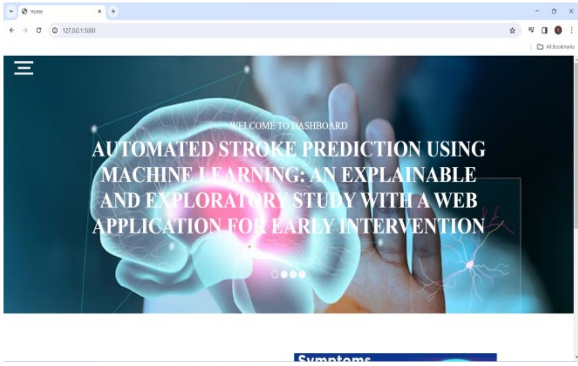
$$\text{F1 Score} = 2 * \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} * 100$$



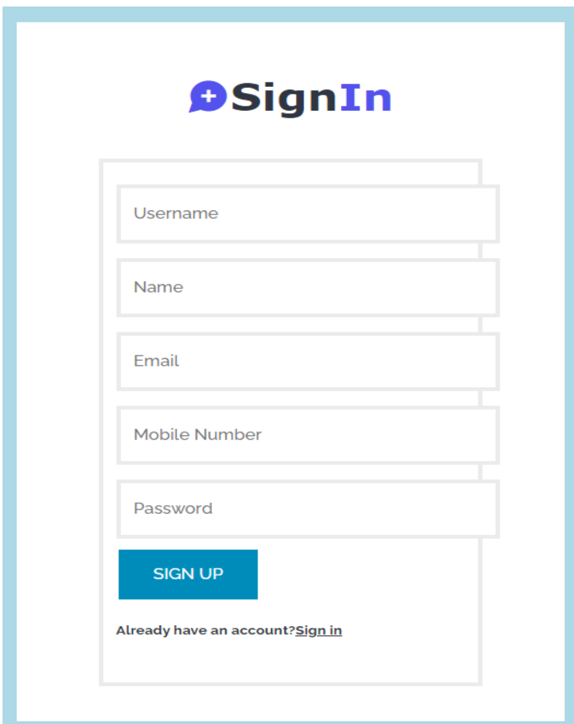
"Fig 14 F1Score"

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 Signin page"

"Fig 18 Login page"

"Fig 19 User input"

"Fig 20 Predict result for given input"

5. CONCLUSION

As part of the project, key measures were taken to "improve the accuracy of stroke prediction by using modern ML methods" to find high-risk individuals. This project ensured that there was a more accurate and fairer combination of strokes and non-stroke cases to maintain the issue of imbalance in the data record lesson and to improve model output. At 99°C, the stack classifier of the extended algorithm performed the best task of predicting strokes. Installed successfully on a user-friendly frontend and processed feature values. This shows both the powerful performance and real usability of health applications. This project makes stroke supply more consistent and effective "by standardizing complex models using global and local explanatory methods. This improves treatment planning to broaden the patient's situation. A major step forward in the field of "Explanatory Artificial Intelligence (XAI) in the medical field is the creation of trustworthy, clear AI systems that provide clear and short explanations. This ensures that the predictive model is responsible and reliable to assess the risk of stroke.

6. FUTURE SCOPE

In the future, researchers might look into a wider range of ML algorithms and methods to make stroke prediction models more accurate and useful in real life. The development of the web tool for early stroke intervention could lead to more business. In the future, more advanced features and functions could be added to make a bigger difference in stroke care and treatment. There are chances to learn more about how to combine global and local explainable methods, especially SHAP. Looking into these methods could help us learn more about how the ML models used in stroke prediction make decisions. Building a complete smart stroke prediction system from start to finish, with mobile apps for both Android and iOS, is one way that could be taken in the future. This growth could make things easier to get to and use. A good direction for future study is to look into demographic factors like age and gender in more depth. Figuring out how these factors affect the risk of stroke in a more complex way could lead to the creation of more accurate and individualized prediction tools.

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.