International Journal of Artificial Intelligence and Mathematical Sciences

December 2025 Vo4. Issue No. 02
ISSN Online: 2958-0153 and Print: 2958-5228
DOI: https://doi.org/10.58921/ijaims.v4i2.174

Brain Stroke Prediction Using Artificial Intelligence and Machine Learning; A Comprehensive Risk Assessment Model

Saima Gul [1], Hafiz Syed Muhammad Kashif [1*], Muhammad Yousuf Tufail [1]

sagul@neduet.edu.pk, engr.muhammadkashifaslam@gmail.com, tufail@neduet.edu.pk

[1] Department of Mathematics, NED University of Eng. & Technology, Karachi, Pakistan 75300

Received: 10-11-2025; **Accepted:** 15-11-2025; **Published**: 30-12-2025

Abstract

Brain stroke refers to a serious condition, happens when the blood flow to the brain is either blocked or affected owing to bleeding due to raptured vessel, in the brain. This lack of blood, thus lack of oxygen supply to brain cells harms brain tissue and leads to quick cell death. Each year, about fifteen million people around the world experienced the stroke. Of these, five million die, and another five million survive but face lasting disabilities. The mortality rate within 28 days after a stroke is about 28%. This study aims to develop a predictive model using artificial intelligence (AI) algorithms for assessing individual stroke risk. It uses a publicly available dataset of about 5,000 records from a website called Kaggle. The predictors include demographic and clinical variables including gender, hypertension, heart disease, smoking status etc. Seven machine learning classifiers, namely Decision Tree, Random Forest Classifier, Support Vector Machine SVM, Naïve Bayes, Logistic Regression, K-Nearest Neighbor KNN, and Gradient Boosting is implemented in Python. These classifiers are compared using different evaluation parameters. The result reveals that the Random Forest classifier consistently outperformed all other models in the evaluation parameters, achieving the highest predictive performance for stroke risk assessment.

Keywords: Stroke prediction, SVM, KNN, Accuracy, Precision, Naïve Bayes, Recall, F1-score, ROC-AUC.

*Email: engr.muhammadkashifaslam@gmail.com

1. INTRODUCTION

Worldwide, stroke takes the second place as the major cause of death after only heart disease [1,2]. The stroke risk factor increases considerably with age and approximately 75% of cases above the age of 65 [3,4]. Most of the deaths due to stroke in the world occur in low-income countries [5]. It is estimated that by the administration of the correct blood pressure, the number of deaths due to strokes could be reduced by nearly 40% [6]. In any case, stroke is still the fifth largest cause of death in the world [7].

Stroke happens when blood flow to the brain is either blocked or interrupted because either a clot has formed in a vessel or a vessel has burst [8,9]. The moment the brain is deprived of blood it loses the supply of oxygen and vital nutrients resulting in brain cell damage or death [10]. The presence of some risk factors makes stroke more likely like; hypertension, smoking, diabetes, hypercholesterolemia, overweight, and inactive lifestyle [11]. The symptoms that require immediate medical attention are sudden numbness, inability to talk or understand, visual disturbances, loss of balance, and very bad headache [12,13]. The technologies of artificial intelligence (AI) have shown great utility and usage in predicting stroke risk, this way allowing early intervention and personalized healthcare strategies which are lifesaving and lessening the global burden of this disabling condition [14,15]. Brain stroke has two types which are [9]:

Ischemic stroke is the most prominent type of stroke, which accounts for around 85% of all diagnosed cases [16]. The process of the ischemic stroke begins with a blood vessel that carries oxygen-rich blood to the brain becoming obstructed. Consequently, the blood supply to that region of the brain suddenly drops or may even completely stop. Usually, such a blockage is due to either a thrombus or an embolus [17,18]. Emergency care is aimed at getting the blood flow back to normal, mainly by administering intravenous thrombolysis using tissue plasminogen activator (tPA) if the patient is treated within four and half hours of the onset of symptoms [19]. Hemorrhagic stroke results from the burst of a damaged blood vessel in the brain, which causes either immediate bleeding within or in the region around the brain. The leaking blood damages brain cells and interferes with regular brain activity. Additionally raises intracranial pressure, lowers nearby cerebral blood flow, and could cause edema and irritation that aggravate the damage [9,20,21].

As shown in Figure 1, one results from a blocked blood vessel while the other comes from a vessel bursting; both have visual distinctions.

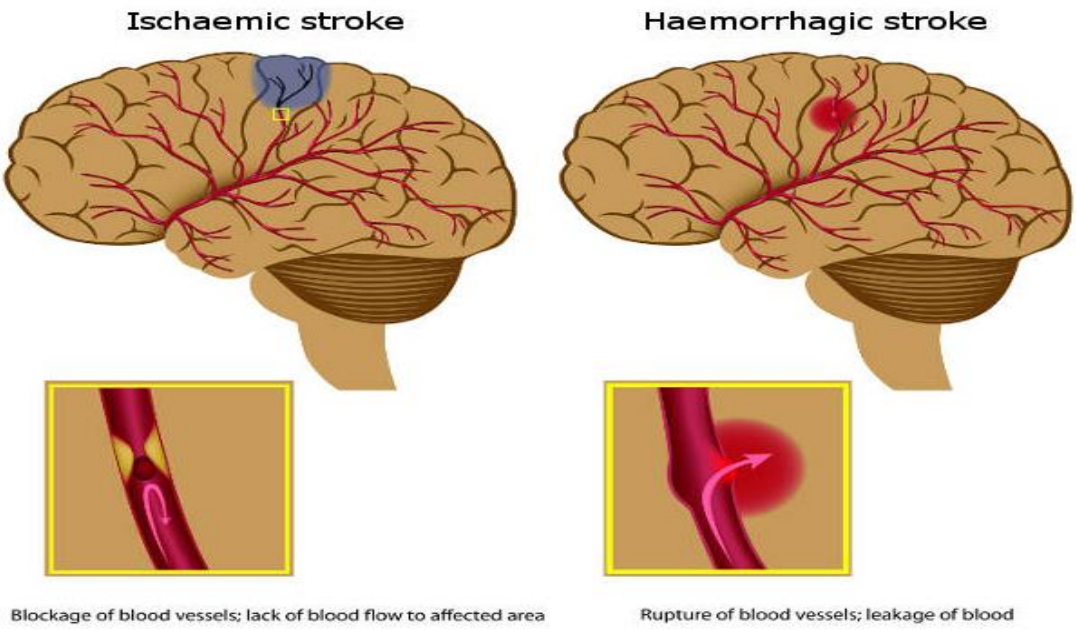

Fig 1. Visualization of Ischaemic and Haemorrhagic Stroke

Research findings indicate that advanced Medical Learning Models, like those found within Artificial Intelligence, can sustain their success rates when diagnosing patients with accuracy greater than vs 80% for experienced clinicians. Advanced ML Models can achieve up to 91 percent on multiple choice charts compared to other forms of clinicians demonstrating the high level of support and improved decision-making capabilities within the medical learning World for Medical Decisions [14,15].

2. LITERATURE REVIEW & RESEARCH GAP

Conventional clinical instruments, such as the Framingham Stroke Risk Profile, have been critically used to guide medical practice; however, they rely heavily on linear regression-based methodologies. Unfortunately, these methodologies are not well suited for handling the complex nonlinear relationships that exist between numerous risk factors. Many of these models are also built for a particular population and do not apply very widely to other ethnic or regional groups, which is yet another constraint [22].

Although their interpretability and simplicity of adoption in clinical practice have made regression-based models still quite popular, several studies have noted their difficulties when working with high-dimensional or nonlinear data [23]. Applied to big datasets with varied clinical characteristics, machine learning models often outperform conventional statistical techniques, as shown by a systematic review [24]. Using a hybrid random forest together with a linear model, one study predicted cardiovascular illnesses and attained 88.7% accuracy utilizing supervised learning algorithms [25]. Ten different ML techniques for early stroke prediction were examined in another study; a weighted voting classifier had the best accuracy and highest AUC [26]. Likewise, other scientists have said that random forests are often the best model for estimating stroke risk [27].

Additional contributions comprise of the application of ML models on stroke data to assess performance disparities among

classifiers, therefore revealing the most efficient model for their dataset [28]. Furthermore, investigations contrasting deep learning approaches with conventional ML methods discovered deep learning does not always provide a major benefit [29]. For example, one analysis employing the International Stroke Trial data found that deep learning techniques did not significantly surpass machine learning classifiers in forecasting ischemic stroke recovery results [30]. Using a stacking technique to stroke dataset, one other study demonstrated great performance in terms of accuracy, AUC, F-measure, accuracy, precision, and recall among other indicators [31].

Researchers are now directed in transparent reporting and prejudice assessment by emerging criteria including TRIPOD-AI and PROBAST-AI [32]. Many ML models produce great discrimination scores like high AUROC values, but when used to outside groups, their calibration usually fails [33]. One writer described calibration as the Achilles heel of predictive analytics since incorrect calibration might deceive doctors and endanger patient safety even if a model has great accuracy [34]. More recently, a research team assessed the calibration and bias of a machine learning model put in use in a major healthcare network [35]. Their results strengthened the idea that wrongly calibrated models can consistently overestimate or underestimate patient risk, hence resulting in incorrect clinical decisions in actual situations [36].

Despite the expansion of machine learning in stroke prediction, there are still several limitations to models being capable of reliably and accurately predicting future occurrence of strokes [14]. Researchers and clinicians continue to push for improved calibration, accuracy, and understanding of AI-based health systems. Many machine learning stroke models suffer from data quality issues due to poor data collection processes (e.g., missing values, inconsistent coding, and changes in clinical practice) that impede long-term performance through model drift, although advancements are occurring at a rapid pace in technology. Additionally, the lack of external validation has decreased the ability of many of these models to perform accurately in different healthcare environments [15]. Addressing these issues requires consistent monitoring, effective governance structures, and the ongoing feedback of clinicians during model development.

In this project we intend to build a stronger, clearer, more reliable stroke risk prediction model than has previously been made available by using machine learning algorithms on an extensive dataset and evaluating our model from multiple dimensions including accuracy, fairness, transparency, and application to real-world data.

3. METHODOLOGY

3.1. Dataset

The study on brain stroke was conducted using a dataset from the Kaggle website containing 4,981 records having 11 features representing demographic, lifestyle and medical data for predicting strokes. Of these features, ten are independent variables while stroke is the dependent variable (or target). The target variable is binary where one indicates the patient has experienced a stroke or has a high chance of having one. A zero indicates there was no history of either event. Table 1 below offers a detailed analysis of each factor such as:

Table 1: Details of dependent and independent variable

Column Name	Type	Values	Definition / Clinical Meaning
(1) gender	Nominal	Female, Male	The patient's biological sex, which can influence their likelihood of experiencing a stroke.
(2) age	Numeric (int/float)	0–100	Age in years (risk increases significantly with age).
(3) hypertension	Binary (0/1)	0 = No, 1 = Yes	Presence of high blood pressure (major stroke risk factor).
(4) heart_disease	Binary (0/1)	0 = No, 1 = Yes	Indicates cardiovascular conditions such as atrial fibrillation.
(5) ever_married	Categorical	Yes, No	Marital status (proxy for social determinants of health).
(6) work_type	Categorical	Private, Self-employed, Govt_job,	Type of employment; linked to lifestyle and healthcare access.
(7) Residence_ type	Categorical	Urban, Rural	Patient's residence location, affecting healthcare accessibility.
(8) avg_glucose_ level	Numeric (float)	50–300 mg/dL	Average blood glucose; high levels indicate diabetes risk.
(9) bmi	Numeric (float)	10–60 kg/m²	Body Mass Index; obesity/underweight linked to stroke risk.
(10) Status of Smoking	Categorical	Never Smoked, Formerly Smoked, Currently Smokes, Unknown	Use of Tobacco history (major modifiable stroke risk factor).
(11) Stroke	Binary (0/1)	0 = No, 1 = Yes	Target variable, Shows if the patient suffered a stroke.

But the dataset has a class imbalance because non-stroke cases are far more than stroke cases. Also, the various data types in the dataset need different preprocessing methods. The pie chart in figure 2 portrays both the classes. Ninety-five percent of the output indicates zero or 'No-Stroke.' Only five percent shows a different result.

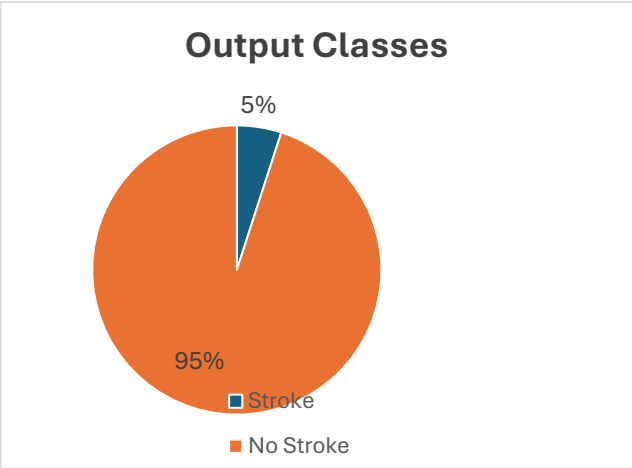

Fig 2. Pie chart shows the imbalance in dataset

3.2. Artificial Intelligence

AI (Artificial Intelligence) is a rapidly evolving area within Computer Science as it relates to developing computer programs that can perform tasks typically performed by humans through intelligence such as: Learning Logic & Reasoning, Understanding the Environment; Making Decisions; Understanding Language; and Solving Complex Problems[37,38,39].

Deeper learning is more sophisticated specialization in machine learning. Special layered neural networks that let the system automatically derive significant characteristics from big and complicated data are employed. Particularly it is good with unstructured data, such as images, audio, video, and text is deep learning. Many times, it outperforms conventional machine learning techniques [40,41]. ANNs and Deep Learning approaches are increasingly being utilized for stroke prediction [42,43]. While deep learning works very well on medical images, time-series data, and electronic health records, ANNs emulate biological neurons and can model highly non-linear relationships [44,45].

A. Logistic Regression

Mostly utilized for classification issues, category prediction this supervised machine learning algorithm is because of its simplicity and interpretability, it is sometimes employed in stroke prediction models as a baseline model [46]. Usually used to model a binary dependent variable with the aid of logistic function, logistic regression is a statistical model [47,48]. Another name for the logistic function is a sigmoid function and is given by:

$$F(x) = \frac{1}{1 + e^{-x}} = \frac{e^{x}}{e^{x} + 1}$$

This feature turns values ranging from very negative to extremely positive into a scale from 0 to 1, therefore assisting the logistic regression model in changing them.

Figure 3 shows the graph between data points on x-axis vs predicted value or probability along y-axis as given below.

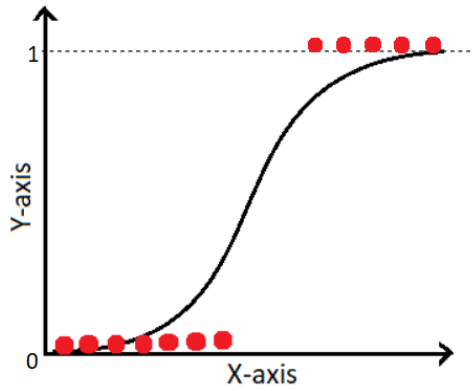

Fig 3. Graph shows the probability 'y' against data.

Simple to execute, works well with smaller datasets, produces probability outputs, that can be used for medical risk prediction and interpret-able coefficients that reveal feature relevance [54,55].

B. Decision Tree

Easy to grasp and often used, decision tree is a ML classifier capable of addressing regression as well as classification problems. Based on various characteristics, it repeatedly divides the data into smaller pieces. The algorithm picks the divide that most clearly divides the data at every stage to create pure and defined groups as feasible [49].

A decision tree resembles a tree. Every internal node depicts a question or condition depending on a feature; each branch depicts the possible outcome of that circumstance; each final node (leaf) provides the result. For classification tasks, the leaf node reveals the predicted class; for regression jobs, it provides a numerical value [50]. The figure 4 below shows the tree structure as;

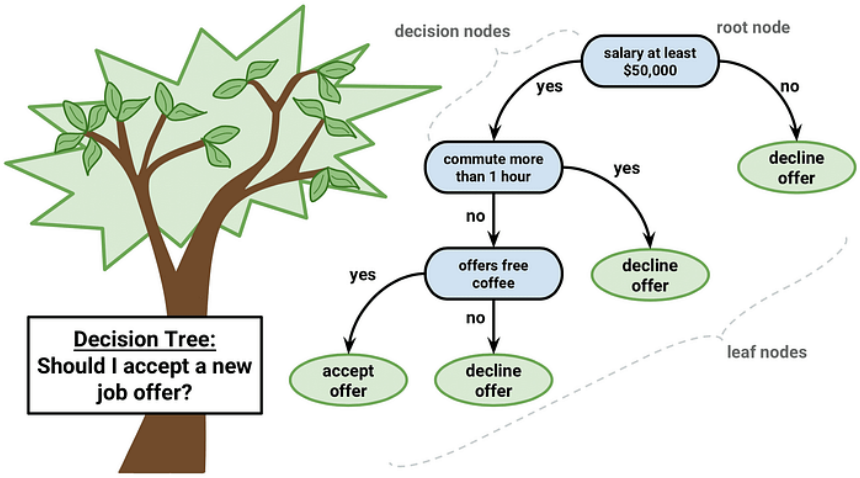

Fig 4. Depiction of decision tree as tree with branches

C. Random Forest

An ensemble machine learning classifier develops multiple decision trees and combines their results to improve prediction accuracy and stability. Instead of relying on a single tree, Random Forest uses bagging, or bootstrap aggregating, where

each tree is trained on a random subset of the data and features. The outcome is determined by majority voting for classification problems or averaging for regression tasks. [51].

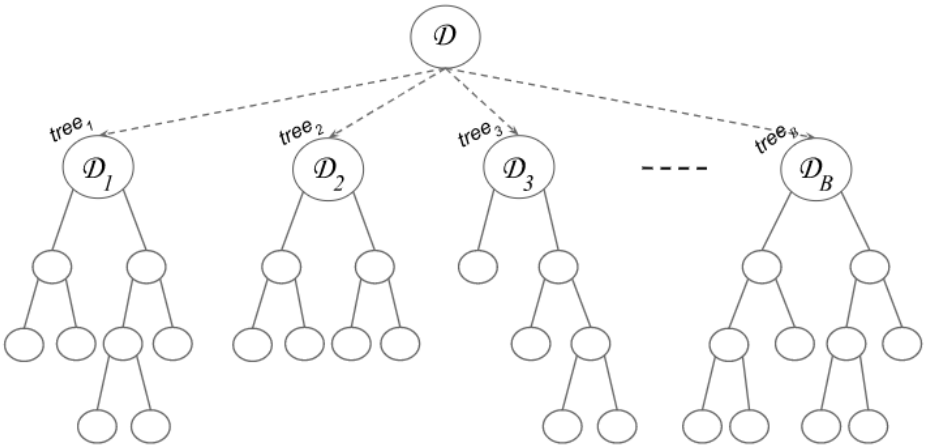

Fig 5. Shows the multiple decision trees of Random Forest

This method enhances generalization on unobserved data and lowers the danger of over-fitting, a often seen problem with single decision trees. Worth in medical prediction models [51]. When we solve regression problems with the Random Forest Algorithm, we use the mean squared error (MSE) to understand how data splits from each node. [52].

D. Gradient Boosting

Gradient boosting is a step-by-step method. In this approach, each new model focuses on correcting the errors made by the previous one. This is different from Random Forest, which builds trees at the same time and in parallel. Gradient boosting improves performance of the model by lowering the loss function through the addition of weak learners, typically shallow decision trees. This method effectively captures complex, non-linear patterns in the data [53].

Gradient boosting gives high predictive accuracy and handles imbalanced datasets well. But it often requires more computing power and careful tuning of parameters. The model can become susceptible to overfitting if parameters like learning rate, tree depth, and estimator count are not carefully tuned [54].

E. Support Vector Machine

Support Primarily intended for classification; the vector machine is a popular supervised learning technique. It can also address regression problems. Its main goal is to find the best boundary that separates different categories of data or hyperplane [55]. The decision line relies heavily on these nearest points, known as support vectors. SVM improves its ability to accurately categorize new and unseen data by increasing this gap, making it reliable for applications that need precision and clear separation.

The SVM model depends on a few key ideas to classify data effectively. At its center is the hyperplane, which serves as the decision boundary and is mathematically given by $wx + b = 0$ in linear cases. The data points closest to this boundary are called support vectors. Even small changes in these points can move the hyperplane. The distance from the hyperplane to these closest points is the margin. SVM aims to maximize this distance to make classification more effective.

F. K-Nearest Neighbours

For both classification and regression applications, KNN is a basic but strong supervised machine learning method. The basic concept centers on resemblance: a fresh data point is given the class most prevalent among its "k" nearest neighbors in the feature space. Usually measures the nearness of points using distance metrics. Non-parametric KNN makes no assumptions about the underlying data distribution, hence it is adaptable and appropriate for a range of intricate datasets [57,58].

The fundamental assumption behind KNN is that neighboring data points are likely to be similar. The following steps the algorithm operates:

Features (attributes) and associated labels (for supervised learning tasks) define the structure of the dataset.

Using a selected distance metric, the algorithm determines the distance from every other point in the dataset when a new point needs to be classified or predicted. The algorithm finds the k nearest neighbors depending on the distances and assigns the most frequent class (for classification) or calculates an average value (for regression) to predict.

G. Naive Bayes

Based on Bayes' theorem of probability, naive Bayes is a straightforward but strong supervised machine learning model. It is termed "nave" since it posits that all characteristics are unrelated of each other, which hardly holds true in actual data but still works surprisingly well in reality. Combining prior probabilities (the overall likelihood) with the conditional probabilities of seeing the given characteristics [59], naive Bayes determines the probability a data point belongs to a certain class.

This classifier presupposes that the characteristics we employ to forecast the target are unrelated and have no bearing on one another. Though in real-world data the independence hypothesis is never true, but often it functions well and hence is termed "Naive".

Bayes Theorem states mathematically that given features vector X=(x1,x2,…,xn) and a class variable y, [60]:

$$P(y|X) = \frac{P(X|y) * P(y)}{P(X)}$$

3.3. Data Preprocessing

The first phase in turning raw data into a tidy, consistent, structured, and practical format for analysis and application of machine learning methods is data preprocessing. Poor preprocessing causes incorrect and biased forecasts [61]; therefore, this phase is rather crucial. The following steps are essential in this process:

A. Data Cleaning

Data cleaning involves detecting, removing and correcting corrupt, inaccurate, inconsistent or irrelevant data from a data-set to raise its quality and dependability and ready it for using algorithms. It comprises a set of activities, addressed below as [62].

B. Handling missing values

First, we check the missing entries, if found, then we either delete row/column with too many missing values, or we we fill missing values using mean, median, mode or other suitable advanced method [63]. In this case of brain stroke data-set, missing values were checked using command ‘ df.isnull().sum() ’ in Python and no missing values were found in any column, as shown in table 2 as:

Table2: Details of null values

```
df.isnull().sum()
✓ 0.3s

gender               0
age                  0
hypertension         0
heart_disease        0
ever_married         0
work_type            0
Residence_type       0
avg_glucose_level    0
bmi                  0
smoking_status       0
stroke               0
dtype: int64
```

C. Data Type Correction

First, it is checked type of entries, broadly numeric and categorical. For this, ‘ df_cat.describe() ’ and ‘ df_num.describe() ’ algorithms applied to find out details of numeric and categorical entries, as shown below in table 3.

Table3: Details of numeric and categorical entries

```
df_cat=df.select_dtypes(np.object_)
df_cat.describe()
✓ 0.0s
```

	gender	ever_married	work_type	Residence_type	smoking_status
count	4981	4981	4981	4981	4981
unique	2	2	4	2	4
top	Female	Yes	Private	Urban	never smoked
freq	2907	3280	2860	2532	1838

```
df_num=df.select_dtypes(np.number)
df_num.describe()
✓ 0.1s
```

	age	hypertension	heart_disease	avg_glucose_level	bmi	stroke
count	4981.000000	4981.000000	4981.000000	4981.000000	4981.000000	4981.000000
mean	43.419859	0.096165	0.055210	105.943562	28.498173	0.049789
std	22.662755	0.294848	0.228412	45.075373	6.790464	0.217531
min	0.080000	0.000000	0.000000	55.120000	14.000000	0.000000
25%	25.000000	0.000000	0.000000	77.230000	23.700000	0.000000
50%	45.000000	0.000000	0.000000	91.850000	28.100000	0.000000
75%	61.000000	0.000000	0.000000	113.860000	32.600000	0.000000

As the data contains both numeric and object type entries, so correction was required to transform all into numeric as algorithms work on numeric values. For this, a technique called ‘LabelEncoder ’ used in Python which converts all the categorical entries such as ‘ yes ’ , ‘ no ’, ‘ male ’ , ‘ female ’ etc into numeric values. ‘ LabelEncoder ’ assigns a specific integer to each different category in the mentioned column [64] as shown here figure 6.

```
le = preprocessing.LabelEncoder()
df_cat['gender'] = le.fit_transform(df.gender.values)
df_cat['ever_married'] = le.fit_transform(df.ever_married.values)
df_cat['work_type'] = le.fit_transform(df.work_type.values)
df_cat['Residence_type'] = le.fit_transform(df.Residence_type.values)
df_cat['smoking_status']= le.fit_transform(df.smoking_status.values)
df_cat
df1 =pd.concat([df_cat,df_num], axis=1)
df1
```

✓ 0.0s Python

	gender	ever_married	work_type	Residence_type	smoking_status	age	hypertension	heart_disease	avg_glucose_level	bmi	strok
0	1	1	1	1	1	67.0	0	1	228.69	36.6	
1	1	1	1	0	2	80.0	0	1	105.92	32.5	
2	0	1	1	1	3	49.0	0	0	171.23	34.4	
3	0	1	2	0	2	79.0	1	0	174.12	24.0	
4	1	1	1	1	1	81.0	0	0	186.21	29.0	
...	...	...	...	...	...	...	...	...	...	...	
4976	1	0	1	0	1	41.0	0	0	70.15	29.8	
4977	1	1	1	1	3	40.0	0	0	191.15	31.1	
4978	0	1	0	0	3	45.0	1	0	95.02	31.8	
4979	1	1	1	0	3	40.0	0	0	83.94	30.0	
4980	0	1	1	1	2	80.0	1	0	83.75	29.1	

Fig 6. Merging of numeric and categorical entries

D. Data Normalization

As the numeric values are very wide ranged, this causes disproportionate influence of features with larger values [65]. So data is transformed to a common scale, typically within a defined range, between zero and one in this case, called data normalization. Normalization is a very crucial pre-processing step for optimal model performance and stability. In python, “ MinMaxScaler( ) ” algorithm is used to perform the process of data normalization.

E. Imbalanced Data and SMOTE

Since dataset is imbalanced, that is one class significantly outnumbers others, here ‘non-stroke’ is predominant over ‘stroke’, as shown and discussed above. This creates severe skews in the class distribution and leads to poor machine learning model performance on minority class. To address this problem, one of the effective techniques is ‘Synthetic Minority Oversampling Technique’ or SMOTE, has been applied on this dataset [66]. In Python, SMOTE can be easily implemented using the imbalanced-learn (imblearn) library, which provides convenient tools for handling class imbalance in machine learning datasets [66].

3.4. Application of algorithms

Understanding the major Python libraries that help with development is important when using machine learning algorithms. Python libraries are collections of pre-written code, including modules, functions, classes, and tools that offer specific features [67,70].

NumPy provides an extensive set of capabilities to enable mathematical calculations to be performed on arrays, and it also supplies efficient means of structuring arrays and matrices to facilitate the efficient retrieval of values stored within them [69]. Developed to enable users to process and analyze large amounts of data, Pandas is an open-source library built on top of NumPy that allows users to easily manipulate structured data such as tables/timeseries. The Scikit-Learn (sklearn) library is an open-source project that provides a simple and effective way to perform data analysis, machine learning, and predictive modelling. The library is built on three main components- NumPy, Scipy, and Matplotlib [68]. Matplotlib offers strong tools to produce legible, enlightening, and aesthetically pleasing charts, matplotlib and seaborn are among the most often used Python data visualization libraries.

Data splitting is an important part before employing predictive classifiers in Python. Typically, the data is separated into training, validation, and testing sets [67]. The training set trains the model by adjusting its parameters. The Scikit-Learn (sklearn) library has the `train_test_split()` function, which simplifies and speeds up this process [69]. In this instance, the data is split into 80% for training and 20 for testing [68,70].

3.5. Type I and Type II Errors

Type I Error (False Positive) occurs when a test incorrectly shows an effect or condition is present when it is not[69,70]. Type II Error (False Negative) occurs when a test fails to identify an effect or condition that is actually present, meaning the null hypothesis is not rejected even though it is false.

Assessing how well a machine learning model performs is an important step. In Python, the Scikit-Learn metrics module offers various tools for this purpose. Accuracy measures the proportion of correct predictions, but it can be misleading with imbalanced datasets. In those cases, metrics like precision, recall, and F1-score give a better evaluation [67].

• Precision is the ratio of correctly predicted positive instances to all predicted positives.

• Recall measures how many actual positives the model correctly identifies.

• F1-score combines precision and recall into a single metric, balancing the two for a more complete assessment.

The confusion matrix breaks predictions into true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). This offers another valuable tool for assessment. Understanding the types of mistakes the model makes is especially important in healthcare and other critical areas. This knowledge helps evaluate its accuracy, summarized in table 4 and formulas in table 5 as;

Table4: Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Table5: List of formula of all the relevant evaluation parameters

Metric	Formula
True positive rate, recall	$\frac{TP}{TP+FN}$
False positive rate	$\frac{FP}{FP+TN}$
Precision	$\frac{TP}{TP+FP}$
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
F-measure	$\frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$

4. RESULTS

By applying and assessing seven commonly used supervised machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, SVM, KNN, and Naïve Bayes, it was observed that Random Forest consistently offered the most reliable and strong performance on this imbalanced dataset. The models were evaluated using key measures such as accuracy, precision, recall, F1-score, and Area Under the ROC Curve (AUC). These measures are important for assessing performance in class-imbalance data.

In relation to all the classifiers tested, Random Forest outperformed each one for accuracy prediction and achieved a better ratio of Precision to Recall. As a result, Random Forest had higher F1 and ROC AUC scores than all the classifiers evaluated in the analysis. Therefore, the performance results indicate that Random Forest is the optimal model for this dataset. The performance scores of all the classifiers across all evaluated metrics can be found in Table 6.

Table5: List of each classifier and their corresponding scores

	Classifier	Accuracy	Precision	Recall	F1 Score	AUC
0	Logistic Regression	0.793559	0.767531	0.842827	0.803419	0.853105
1	Decision Tree	0.912355	0.903093	0.924051	0.913452	0.912342
2	Random Forest	0.954065	0.946114	0.963080	0.954522	0.991774
3	Gradient Boosting	0.897571	0.885481	0.913502	0.899273	0.969824
4	Support Vector Machine	0.834741	0.794254	0.904008	0.845585	0.902877
5	K-Nearest Neighbors	0.909187	0.853370	0.988397	0.915934	0.961141
6	Naive Bayes	0.771383	0.750730	0.813291	0.780759	0.837857

5. CONCLUSION

The focus of this study was to determine how well the seven most widely used classifiers for supervised machine learning performed: Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, SVM, KNN, and Naïve Bayes, according to several performance metrics (accuracy, precision, recall, F-1 Score, AUC). Based on all of these performance metrics Random Forest constantly outperformed the other methods. The fact that Random Forest is capable of handling non-linearity, reduces overfitting through ensemble methods as well as having the ability to provide balanced classifications was the reason why Random Forest was selected for use with this dataset. The findings of this study suggest that the use of ensemble approaches, such as Random Forest, is ideal for classifying problems associated with imbalanced data.

6. FUTURE WORKS

Traditional machine learning methodologies can struggle with ambiguous or borderline cases that occur with imbalanced datasets. The application of a fuzzy logic approach such as Fuzzy Rule-Based Classifiers, Adaptive Neuro-Fuzzy Inference Systems (ANFIS) or Fuzzy Ensemble Models may provide added support to improve decision-making techniques since they provide degrees of membership rather than strict classification labels.

As researchers continue to develop larger and more complex data sets, Quantum Computing provides researchers a way to drastically improve their classification accuracy and therefore, model training speed. Future research could focus on the development of Quantum Support Vector Machines (QSVM), Quantum Neural Networks (QNN), and Quantum Enhanced Ensemble Methods. By combining both Classical and Quantum Learning methods this transition may also provide researchers with a means to reduce costs, enhance learning in high dimensional spaces and, produce results that are superior to both the Classical.

References

[1] Mackay, J., & Mensah, G. A. (2004). The atlas of heart disease and stroke. World Health Organization.

[2] Pandian, J. D., Gall, S. L., Kate, M. P., Silva, G. S., Akinyemi, R. O., Ovbiagele, B. I., ... & Thrift, A. G. (2018). Prevention of stroke: a global perspective. The Lancet, 392(10154), 1269-1278.

[3] George, M. G., Tong, X., & Bowman, B. A. (2017). Prevalence of cardiovascular risk factors and strokes in younger adults. JAMA neurology, 74(6), 695-703.

[4] Sacco, R. L., Benjamin, E. J., Broderick, J. P., Dyken, M., Easton, J. D., Feinberg, W. M., ... & Wolf, P. A. (1997). Risk factors. Stroke, 28(7), 1507-1517.

[5] Strong, K., Mathers, C., & Bonita, R. (2007). Preventing stroke: saving lives around the world. The Lancet Neurology, 6(2), 182-187.

[6] Lawes, C. M., Bennett, D. A., Feigin, V. L., & Rodgers, A. (2004). Blood pressure and stroke: an overview of published reviews. Stroke, 35(3), 776-785.

[7] Mukherjee, D., & Patil, C. G. (2011). Epidemiology and the global burden of stroke. World neurosurgery, 76(6), S85-S90.

[8] Schroeder, E. B., Rosamond, W. D., Morris, D. L., Evenson, K. R., & Hinn, A. R. (2000). Determinants of use of emergency medical services in a population with stroke symptoms: the Second Delay in Accessing Stroke Healthcare (DASH II) Study. Stroke, 31(11), 2591-2596.

[9] Sirsat, M. S., Fermé, E., & Câmara, J. (2020). Machine learning for brain stroke: a review. Journal of Stroke and Cerebrovascular Diseases, 29(10), 105162.

[10] Amarenco, P., Bogousslavsky, J., Caplan, L. R., Donnan, G. A., & Hennerici, M. G. (2009). Classification of stroke subtypes. Cerebrovascular diseases, 27(5), 493-501.

[11] Boehme, A. K., Esenwa, C., & Elkind, M. S. (2017). Stroke risk factors, genetics, and prevention. Circulation research, 120(3), 472-495.

[12] Aigner, A., Grittner, U., Rolfs, A., Norrving, B., Siegerink, B., & Busch, M. A. (2017). Contribution of established stroke risk factors to the burden of stroke in young adults. Stroke, 48(7), 1744-1751.

[13] Robertson, I. H., & Murre, J. M. (1999). Rehabilitation of brain damage: brain plasticity and principles of guided recovery. Psychological bulletin, 125(5), 544.

[14] Bonkhoff, A. K., & Grefkes, C. (2022). Precision medicine in stroke: towards personalized outcome predictions using artificial intelligence. Brain, 145(2), 457-475.

[15] Mainali, S., Darsie, M. E., & Smetana, K. S. (2021). Machine learning in action: stroke diagnosis and outcome prediction. Frontiers in neurology, 12, 734345.

[16] Feske, S. K. (2021). Ischemic stroke. The American journal of medicine, 134(12), 1457-1464.

[17] Herpich, F., & Rincon, F. (2020). Management of acute ischemic stroke. Critical care medicine, 48(11), 1654-1663.

[18] Caplan, L. R. (1993). Brain embolism, revisited. Neurology, 43(7), 1281-1281.

[19] National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. (1995). Tissue plasminogen activator for acute ischemic stroke. New England Journal of Medicine, 333(24), 1581-1588.

[20] Montaño, A., Hanley, D. F., & Hemphill III, J. C. (2021). Hemorrhagic stroke. Handbook of clinical neurology, 176, 229-248.

[21] Teunissen, L. L., Rinkel, G. J., Algra, A., & Van Gijn, J. (1996). Risk factors for subarachnoid hemorrhage: a systematic review. Stroke, 27(3), 544-549.

[22] Wolf, P. A., D'Agostino, R. B., Belanger, A. J., & Kannel, W. B. (1991). Probability of stroke: a risk profile from the Framingham Study. Stroke, 22(3), 312-318.

[23] Makke, N., & Chawla, S. (2024). Interpretable scientific discovery with symbolic regression: a review. Artificial Intelligence Review, 57(1), 2.

[24] Rajula, H. S. R., Verlato, G., Manchia, M., Antonucci, N., & Fanos, V. (2020). Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. Medicina, 56(9), 455.

[25] Rimal, Y., Sharma, N., Paudel, S., Alsadoon, A., Koirala, M. P., & Gill, S. (2025). Comparative analysis of heart disease prediction using logistic regression, SVM, KNN, and random forest with cross-validation for improved accuracy. Scientific Reports, 15(1), 13444.

[26] Emon, M. U., Keya, M. S., Meghla, T. I., Rahman, M. M., Al Mamun, M. S., & Kaiser, M. S. (2020, November). Performance analysis of machine learning approaches in stroke prediction. In 2020 4th international conference on electronics, communication and aerospace technology (ICECA) (pp. 1464-1469). IEEE.

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

[28] Biswas, N., Uddin, K. M. M., Rikta, S. T., & Dey, S. K. (2022). A comparative analysis of machine learning classifiers for stroke prediction: A predictive analytics approach. Healthcare Analytics, 2, 100116.

[29] Nielsen, A., Hansen, M. B., Tietze, A., & Mouridsen, K. (2018). Prediction of tissue outcome and assessment of treatment effect in acute ischemic stroke using deep learning. Stroke, 49(6), 1394-1401.

[30] Hung, C. Y., Chen, W. C., Lai, P. T., Lin, C. H., & Lee, C. C. (2017, July). Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database. In 2017 39th annual international conference of the IEEE engineering in medicine and biology society (EMBC) (pp. 3110-3113). IEEE.

[31] Mondal, S., Ghosh, S., & Nag, A. (2024). Brain stroke prediction model based on boosting and stacking ensemble approach. International Journal of Information Technology, 16(1), 437-446.

[32] Collins, G. S., Dhiman, P., Navarro, C. L. A., Ma, J., Hooft, L., Reitsma, J. B., ... & Moons, K. G. (2021). Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ open, 11(7), e048008.

[33] Bella, A., Ferri, C., Hernández-Orallo, J., & Ramírez-Quintana, M. J. (2010). Calibration of machine learning models. In Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques (pp. 128-146). IGI Global Scientific Publishing.

[34] Zimmerman, N., Presto, A. A., Kumar, S. P., Gu, J., Hauryliuk, A., Robinson, E. S., & Robinson, A. L. (2018). A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. Atmospheric Measurement Techniques, 11(1), 291-313.

[35] Huang, J., Galal, G., Etemadi, M., & Vaidyanathan, M. (2022). Evaluation and mitigation of racial bias in clinical machine learning models: scoping review. JMIR medical informatics, 10(5), e36388.

[36] Straw, I., & Wu, H. (2022). Investigating for bias in healthcare algorithms: a sex-stratified analysis of supervised machine learning models in liver disease prediction. BMJ health & care informatics, 29(1), e100457.

[37] Holmes, J., Sacchi, L., & Bellazzi, R. (2004). Artificial intelligence in medicine. Ann R Coll Surg Engl, 86(86), 334-8.

[38] Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. Wiley interdisciplinary reviews: data mining and knowledge discovery, 9(4), e1312.

[39] Mukhamediev, R. I., Popova, Y., Kuchin, Y., Zaitseva, E., Kalimoldayev, A., Symagulov, A., ... & Yelis, M. (2022). Review of artificial intelligence and machine learning technologies: classification, restrictions, opportunities and challenges. Mathematics, 10(15), 2552.

[40] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

[41] Pandey, D., Niwaria, K., & Chourasia, B. (2019). Machine learning algorithms: a review. Mach. Learn, 6(2).

[42] Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. SN computer science, 2(3), 160.

[43] Qamar, R., & Zardari, B. A. (2023). Artificial neural networks: An overview. Mesopotamian Journal of Computer Science, 2023, 124-133.

[44] Abdou, M. A. (2022). Literature review: Efficient deep neural networks techniques for medical image analysis. Neural Computing and Applications, 34(8), 5791-5812.

[45] Banerjee, I., Ling, Y., Chen, M. C., Hasan, S. A., Langlotz, C. P., Moradzadeh, N., ... & Lungren, M. P. (2019). Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification. Artificial intelligence in medicine, 97, 79-88.

[46] Ng, A., & Jordan, M. (2001). On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. Advances in neural information processing systems, 14.

[47] Schein, A. I., & Ungar, L. H. (2007). Active learning for logistic regression: an evaluation. Machine Learning, 68(3), 235-265.

[48] LaValley, M. P. (2008). Logistic regression. Circulation, 117(18), 2395-2399.

[49] Safavian, S. R., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. IEEE transactions on systems, man, and cybernetics, 21(3), 660-674.

[50] Priyanka, & Kumar, D. (2020). Decision tree classifier: a detailed survey. International Journal of Information and Decision Sciences, 12(3), 246-269.

[51] Pal, M. (2005). Random forest classifier for remote sensing classification. International journal of remote sensing, 26(1), 217-222.

[52] Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS journal of photogrammetry and remote sensing, 114, 24-31.

[53] Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review, 54(3), 1937-1967.

[54] Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. Frontiers in neurorobotics, 7, 21.

[55] Suthaharan, S. (2016). Support vector machine. In Machine learning models and algorithms for big data classification: thinking with examples for effective learning (pp. 207-235). Boston, MA: Springer US.

[56] Fung, G., & Mangasarian, O. L. (2001, August). Proximal support vector machine classifiers. In Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 77-86).

[57] Peterson, L. E. (2009). K-nearest neighbor. Scholarpedia, 4(2), 1883.

[58] Liao, Y., & Vemuri, V. R. (2002). Use of k-nearest neighbor classifier for intrusion detection. Computers & security, 21(5), 439-448.

[59] Patil, T. R., & Sherekar, S. S. (2013). Performance analysis of Naive Bayes and J48 classification algorithm for data classification. International journal of computer science and applications, 6(2), 256-261.

[60] Dai, W., Xue, G. R., Yang, Q., & Yu, Y. (2007, July). Transferring naive bayes classifiers for text classification. In AAAI (Vol. 7, pp. 540-545).

[61] Kamiran, F., & Calders, T. (2012). Data preprocessing techniques for classification without discrimination. Knowledge and information systems, 33(1), 1-33.

[62] Chu, X., Ilyas, I. F., Krishnan, S., & Wang, J. (2016, June). Data cleaning: Overview and emerging challenges. In Proceedings of the 2016 international conference on management of data (pp. 2201-2206).

[63] Saar-Tsechansky, M., & Provost, F. (2007). Handling missing values when applying classification models.

[64] Foken, T., Leuning, R., Oncley, S. R., Mauder, M., & Aubinet, M. (2011). Corrections and data quality control. In Eddy covariance: a practical guide to measurement and data analysis (pp. 85-131). Dordrecht: Springer Netherlands.

[65] Singh, D., & Singh, B. (2020). Investigating the impact of data normalization on classification performance. Applied Soft Computing, 97, 105524.

[66] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.

[67] Raschka, S., Patterson, J., & Nolet, C. (2020). Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence. Information, 11(4), 193.

[68] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825-2830.

[69] Müller, A. C., & Guido, S. (2016). Introduction to machine learning with Python: a guide for data scientists. " O'Reilly Media, Inc.".

[70] Van Rossum, G., & Drake Jr, F. L. (1995). Python tutorial (Vol. 620). Amsterdam, The Netherlands: Centrum voor Wiskunde en Informatica.