# A Review on Predicting Brain Stroke using Machine Learning

**Shehzada Mushtaq**
Department of Computer Science and Engineering
Chandigarh University
Gharuan, Mohali, India
shehzadamushtaq12@gmail.com

**Kamaljit Singh Saini**
Department of Computer Science and Engineering
Chandigarh University
Gharuan, Mohali, India
kamaljit.cse@cumail.in

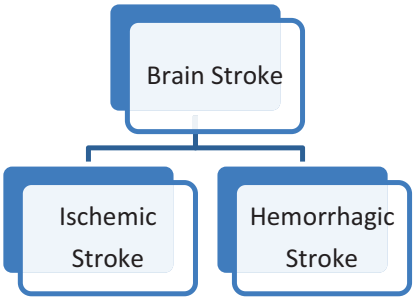
***Abstract*—At present, healthcare is one of the biggest concerns in the world. Brain Stroke is the leading cause of death worldwide. Prediction and detection of the occurrences of a brain stroke at the early stages is a valuable work in the medical field. There are several factors which are responsible for the brain stroke such as BMI (Body Mass Index); Age; Sex; Family Background; Gender; smoking status; hypertension, etc. In healthcare, a lot of research work has been done on the prediction of heart disease, but relatively little attention is paid on predicting a brain stroke. The main aim of this study is to review different research articles published earlier and to choose the best machine learning techniques for the prediction of brain stroke for our future work. After reviewing the different machine learning methods utilized for stroke predictions and after taking into account the previously published studies, it has been discovered that death rate and functional results are the expected outcomes for the majority of the research work done. Support Vector Machine, Stacking, Decision Tree, Weighted Voting, Random Forest, Neural Networks and Naive Bayes were the most frequently employed methods.**

***Keywords*—Machine Learning, Classification Algorithms, Support Vector Machine (SVM), Supervised learning, Prediction, Confusion Matrix, Accuracy, Precision, Random Forest, Naive Bayes**

## I. Introduction

A stroke is a potentially fatal condition in which blood vessels in the brain rupture and the important nutrients and oxygen which the blood carries doesn't reach to the brain tissues in different part of the brain causing damage to brain cells [1]. Stroke has a high mortality rate. One side of the body feels weak or numb, having trouble in speaking or understanding others, having trouble in seeing and having a terrible headache are some of the symptoms of a stroke [2]. Stroke is named for the manner in which it incapacitates individuals. To avoid more damage to the afflicted part of the brain and serious issues in various parts of the body, early detection and effective measures are required. It affects more than 16 million people globally each year and more than five million people lost their lives, and around five million people became permanently inactive, putting a burden on their families and relatives [3]. Stroke was previously seen mostly in peoples above the age of 40, but it is now more common in adults as well. High blood pressure is the major cause of the brain stroke [4]. Apart from high BP (Blood Pressure) there are several other factors which are also responsible for the stroke that includes Family background, Gender, Type of work, Hypertension, Marriage status, etc. Stroke, on the other side, impacts around 8% of kids with sickle cell disease [3]. There are basically two major kind of brain stroke (Figure 1).

- Ischemic stroke
- Hemorrhagic stroke




### A. Ischemic Stroke

It is a type of stroke in which the blood clot in blood artery in the brain prevents blood from carrying important nutrients and oxygen to the brain cells [5]. The fatty deposit present on the inner lining of a blood artery in the brain that is Atherosclerosis is the main cause of the blood clotting. The fatty deposit causes fragment, and obstruct blood flow to the brain. Similar to a heart attack, the concept is that a blood clot prevents blood from flowing to a certain area of your heart. The blood clot that results in an ischemic stroke, as opposed to a Transient Ischemic Attack (TIA) won't go away on its own. Depending on how early you arrive at a hospital, you will receive different ischemic stroke therapies. They also rely on the details of your medical background. Once you get medical assistance for this kind of stroke within three hours, your doctor might be able to provide tissue plasminogen activator to you. The clot can be physically removed by medical procedures or treated with clot-busting drugs administered directly to the brain. This is the most frequent type of stroke, accounting for around 87% of all brain strokes [1]. Figure 2 shows top 10 reasons of mortality worldwide [6].

### B. Hemorrhagic Stroke

This type of stroke occurs when a blood artery ruptures in the brain and bleeding occurs inside the brain [7]. These strokes are classified into three categories: The first one is an Aneurysm that bursts and then forces a piece of the weak blood artery to expand outward. The second one includes irregularly shaped blood arteries and is known as an arteriovenous malformation. These blood arteries might rupture and cause hemorrhagic strokes.

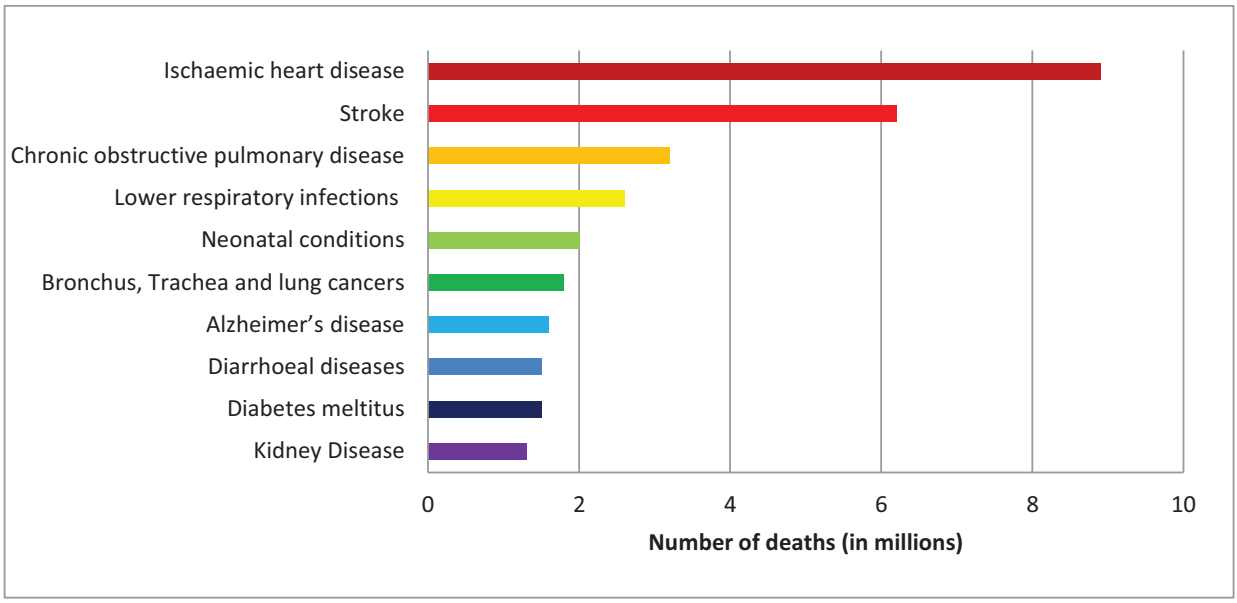



Fig. 1. Top 10 reasons of mortality worldwide [6], recorded in 2019

Another factor is the excessively high blood pressure that may damage the brain's tiny blood artery, which might results in a brain hemorrhage. Treatments for hemorrhagic stroke focus on controlling blood bleeding in the brain and minimizing the negative effects of the brain hemorrhage. A rise in intracranial pressure is a potential side effect. Surgical practices include coiling or clipping. These are intended to stop additional bleeding from the blood vessels in the brain.

According to a survey done by CDC (Center for Disease Control and Prevention) in the USA the stroke occurs in more than 79 lakh 5 thousand individuals and among them 61 lakh of individuals experience a stroke for the first time and rest of individuals had already have a previous stroke, additionally black people have nearly a twofold increased risk of stroke compared to white people [3].

At present, as medical technology develops, it is now feasible to predict the occurrence of any disease using the machine learning techniques [8]. Machine Learning (ML) Techniques are helpful in building correct estimates and for providing proper analysis. The majority of past work has concentrated on cardiac stroke detection. Brain stroke disease prediction has received very little importance. The goal of this review paper is to find out the most appropriates machine learning approaches for stroke prediction, which will also serve as a better understanding and the solution of the issue.

## II. Machine Learning

Machine Learning is a sub-field of Computer Science more specifically Artificial Intelligence. It makes it possible for computers to effectively learn from historical data and experience without explicit instructions [9]. There are 3 main types of Machine Learning which are Unsupervised Learning, Supervised and Reinforcement learning and will be discussed in the next sub-section. Machine Learning Algorithms are used for different application which includes Speech Recognition, Prediction, Computer Vision, Robotics and Data Classification.

### *A. Types and Importance of Machine Learning*

Three different types of machine learning are Supervised Learning, Unsupervised Learning, and Reinforcement Learning. In this section we will discuss them briefly.

#### *1) Supervised Learning*

In Supervised Learning the data used to train the model is already labeled and based on that labeled data the machine predicts the future outcomes [10]. Supervised Learning is further divided into two categories which are Regression and Classification.

- Regression: This is a type of Supervised Learning and is used when some relationship exists between the input variable and the output variable and is used when the expected result is a continuous variable [11]. Linear regression, regression trees, polynomial regression and non-linear regression are some types of it.

- Classification: This type of Supervised Learning is used when some relationship exists between the input features and the target feature and is used when the expected result is a categorical variable which means that there are classes like true-false, male-female, and apple-banana. Random forest, support vector machine, decision tree and naive bayes are some of the types of classification algorithm [12]. This Review paper is mostly based on the study of supervised learning techniques for the prediction of brain stroke.

#### *2) Unsupervised Learning*

In case of unsupervised learning the data that we provide to the machine for the training is unlabeled and the machine itself find the hidden pattern with the help of Artificial Intelligence algorithms and predict the outcome [13]. Clustering and association are the types of unsupervised algorithm.

- Clustering: Clustering is a supervised Machine Learning technique that groups unlabeled data into clusters. This technique groups similar data into one cluster and non- similar to the other cluster hence named clustering [14].

*3) Reinforcement Learning*

It is a type of learning algorithm in which an agent requires a series of decisions to reach a final destination [13]. Artificial Agent interacts with the environment and for every right and the wrong actions it performs, it will get either reward or penalty based on the action it does. The main goal is to maximize the reward.

Despite the fact that machine learning is always changing with new technology, it is still employed in a variety of sectors. Machine learning is critical because it gives organizations knowledge of customer behavior trends and operational business patterns, as well as helping in the creation of new good and services. With the help of the machine learning algorithms, we can easily forecast the future outcome and we can find the hidden pattern from a large volume of data [15]. Today, many of the companies like Facebook, Google and Tesla are using machine learning. Also, it has a lot of real world applications in our life form banking to Healthcare and from Personal Assistance to Product Recommendation.

## III. Machine Learning for Prediction of the Brain Stroke

This section analyses and reviews previously published papers related to studies on stroke prediction using the various machine learning algorithms. They are discussed below:

Heo et al. [16] used different Machine Learning Algorithms, depending on particular characteristics, Deep Neural Network (DNN), Random Forest and the Logistic Regression for the prediction of Stroke. We learned from this paper that deep neural networks (DNNs) are commonly utilized for acute or ischemia stroke patients, and that they have an effect on long-term forecasting. The AUC (Area Under Curve) of the DNN model with respect to input was 88.8 percent, which was higher than the other models. Automated and more precise calculations are used to enhance the model.

Jenna et al. [17] applied Support Vector Machine with some kernel for the early prediction and detection of the brain stroke. They used dataset from the International Stroke trial Database for their research. 350 inputs were used for the prediction after pre-processing to eliminate duplicate and conflicting data. It was performed on MATLAB, and the accuracy achieved was 91 percent.

Alotaibi [18] showed the construction of a machine learning system for the forecasting of the heart stroke. The effectiveness of several machine learning models, such as, Naive Bayes, Support Vector Machine (SVM), and Decision Tree were examined to build the model. The algorithms used had a maximum accuracy of about 93 percent, which was extremely very good.

Sailasya and Kumari [19] in their research used stroke dataset taken from a publicly open website and initially done pre- processing on the dataset and balanced the data using the under-sampling technique. After preprocessing they applied six different ML classification algorithms on the dataset and compared all of them and found out that the Naive Bayes algorithm accuracy was the better than other algorithm with good precision, F1-score and recall score and after all this they had created a HTML page to make it user friendly.

Sharma et al. [20], in their research used the dataset that was publicly available on the kaggle repository and they implemented this in the python using 5 different classification algorithm and used Feature Selection with Ranker Method for better prediction and find out that the Random Forest out perform all the other models and give better result with high accuracy. Tazin et al. [21], conducted research on stroke pre- diction using 4 different classification algorithms and used the dataset from the kaggle repository. After handling missing values and doing over-sampling the different classification models were compared and founded that Random Forest has the best accuracy of 96 percent among the other 3 models.

Harshita et al. [22], conducted research on the stroke dataset taken from kaggle for stroke prediction and used 5 different machine learning classification algorithms and did a comparison of them. They found out that Logistic Regression and Random Forest have the highest accuracy of 95.5% but they had chosen the Random Forest algorithm for their application as it has a lower false negative rate and high accuracy. Dev et al. [23], in their research uses PCA and different statistical techniques and founded that heart disease, hypertension, age and average glucose level as the most essential attributes for the prediction of the stroke disease. The dataset taken was highly imbalanced so they have used sub-sampling techniques to balance it. After this, they had applied various classification algorithms and founded that Neural Network achieved the highest accuracy of 78 % with a low miss rate than others.

Shoily et al. [24], evaluated various algorithms such as Naive Bayes, J48, K-Nearest Neighbor and RF (Random Forest) and discovered that Naive Bayes has higher precision. Various medical records were examined in order for the dataset to be compiled. It was also cross-referenced by healthcare specialists and utilized with the help of Weka tool. The proposed model assists patients in determining whether or not they are at risk of having a stroke. K-NN, J48, Naive Bayes, and Random Forest were among the models that were trained. To validate the models, precision and accuracy were observed. Machine learning models are trained using the dataset which was taken from open source repository kaggle.

For the classification of stroke risk levels different algorithms are used such as Naive Bayes, Bagged Decision trees, Decision Trees, Logistic Regression, Voting, Neural Networks, Bayesian Networks, Random Forest and boosting technique with Decision trees (DT). In this research work they founded that the RF classifier has the best precision of 97.33 percent while the boosting technique with DT has the best recall of 99.94 % [25].

The researcher for their work obtained stroke data from Sugam Hospital in Kumbakonam, Tamil Nadu, India and categories the kind of stroke using different data mining and machine learning methods. ANN (Artificial Neural Network)

trained with the approach of stochastic gradient descent beat other techniques, with good classification results of having accuracy more than 95 percent, whereas the divisions of SVM and ensemble offered accuracy of 91 percent [26].

In order to forecast stroke, Hager et al. [27] applied 4 different types of supervised classification models which are decision trees, SVM, logistic regression and RF. To obtain the result, they used Hyper-parameter tuning and cross-validation on different classification algorithms. They then assessed the effectiveness of these four models, founded that Random Forest had the best accuracy rate of 90% compared to other. In their study on stroke, Minhaz et al. [28] collected stroke data from various Hospitals of Bangladesh. Ten training algorithms are used after the data preprocessing. The weighted voting classifier is then used to improve the efficiency of each classifier. The best model is then founded using a weighted vote classifier after every model has been optimized. According to the study's findings, the weighted voting classifier has the highest accuracy of 97%.

Kaur et al. [29], for their research collected data using EEG (Electroencephalography) technique and then ported the bio-signal data to the server. After this, the preprocessing of the data was done. For the prediction of the brain stroke various Deep Learning Techniques were applied on the data. The Gated Recurrent Unit Network (GRU) technique of deep learning performs better with 95.6% accuracy followed by Bidirectional Long Short-Term Memory (BILSTM) having accuracy of 91%.

Mustafa et al. [30], proposed a prediction model for brain stroke using various Machine Learning techniques. The dataset was taken from kaggle and the required preprocessing was done on it. The data was then balanced using SMOTE technique and the best features are considered using the Hyper-parameter Tuning Technique. Then 5 different Machine Learning classifiers were used as a Base Learner for building Stacking Model in level 1 and then in level 2 the Random Forest was used as a Meta-Learner for achieving more accuracy. The Accuracy obtained using this Stacking ensemble classifier was 97%.

Kumar [31] has taken dataset from kaggle which consists of eleven features. The dataset taken was imbalanced they balanced it using the SMOTE technique. In their research XGBoost model was used and the accuracy achieved was 98%.By using the over-sampling technique instead of under-sampling technique improves the result by 46%.

Gupta et al. [32] conducted research on stroke prediction. The dataset was taken from the kaggle repository and preprocessing was done on the dataset. After this imbalanced dataset was balanced using the Oversampling SMOTE Technique. The dataset was then divided in ratio 80:20 and after this seven different Machine Learning Algorithms were trained on this data. Finally, it was concluded that the AdaBoost, Random Forest and XGBoost had the least value of incorrect prediction with the highest accuracy in case of Random Forest of 97%.

Zubaidi et al. [33] used different supervised algorithm for the forecasting of stroke. After pre-processing of the dataset, the dataset was balanced using the oversampling technique. The dataset was then divided in ratio 80:20 as the training and testing set. Model was then trained using five different classification algorithms. The highest accuracy was achieved in case of the Random Forest classifier of 94.7%.

Almadani et al. [34] build a stroke prediction model using three different classification algorithms. The dataset for this research was obtained from King Abdul-Aziz Medical City (KAMC), Saudi Arabia. After different preprocessing stages the dataset was trained using the machine learning algorithms on the WEKA environment. On comparison they found that the J84 method using PCA achieved the highest accuracy of 95.5%. Monteiro et al. [35], created a model to predict the functional diagnosis of a stroke. The dataset taken for this research consist the record of 541 patients. The features with missing values were removed. After preprocessing there were 152 features and 425 patients record left. 5 different machine learning algorithms were used to construct the model. The performance of the models was evaluated using the area under the curve (AUC) metrics, which is more than 90%.

Rahman et al. [36] proposed a machine learning classification and deep learning algorithm based approaches for predicting the stroke. The stroke dataset was taken from the kaggle repository for research work. The results showed that the machine learning approaches, especially the random forest outperforms deep neural network technique by achieving an accuracy of 99%. Dritsas et al. [37] used the machine learning based approaches for the prediction of the brain stroke using stroke dataset obtained from kaggle. The research was mainly focused on predicting brain stroke peoples who are over eighteen years. After preprocessing the model was trained on 9 different algorithms. The model was evaluated on the basis of metrics such as the Accuracy, Precision, Area under the Curve (AUC), F-measure and Recall. It was found that the stacking method performs better than other Machine learning algorithms with AUC of 98.9 percent, recall, precision and F-measure of 97.4% and accuracy 98%.

A summary of all the research work studied on the brain stroke prediction is summarized in Table I.

## IV. Results and Analysis

A total of 22 studies met the inclusion criteria and were included in this review. The studies varied in terms of the type of predictive model used, the population studied, and the outcome measures assessed. The findings of this review suggest that predictive models can be effective in identifying individuals at risk of experiencing a stroke. Out of 22 studies, 18 reported that their predictive model had good or excellent discrimination, indicating that it was able to accurately distinguish between individuals who were at high risk of experiencing a stroke and those who were not. The majority of the studies which are included in this paper used a combination of clinical, demographic, and imaging data to develop their predictive models. For example, some studies used information on age, residence, BMI, work type, gender, and smoking status to develop their models. Other studies incorporated imaging data, such as MRI or CT scans, to identify individuals with evidence of cerebrovascular disease. The results of different studies based on the same dataset i.e., Healthcare Kaggle dataset on the stroke is summarized in Table II with different performance metrics. The visualized comparative analysis of different studies on Healthcare Kaggle stroke dataset is shown in Fig. 3.

TABLE I. COMPARISON OF VARIOUS STUDIES ON BRAIN STROKE PREDICTION USING ML MODELS

| Study | Algorithms used | Dataset | Result |
|---|---|---|---|
| Heo et al. [16] | Random forest (RF), Logistic Regression (RF) Deep Neural Network (DNN) and ASTRAL | Private Dataset with records of 3522 patients | Accuracy 95% by DNN |
| Jeena et al. [17] | Support Vector Machine (SVM) | Dataset from International Stroke Trail Database | Accuracy 91% |
| Sailasya et al. [19]] | LR, SVM, RF, K-Nearest Neighbor (K-NN) and Naive Bayes (NB) | Healthcare Stroke Data from Kaggle Repository | Accuracy 82% by Naive Bayes |
| Sharama et al. [20] | LR, Decision Tree (DT), RF, JRip and Multi-Layered Perceptron (MLP) | Stroke Prediction Dataset from Kaggle Repository | Accuracy 98.94% by RF |
| Tazin et al. [21] | LR, DT, RF and Voting Classifier | Stroke Prediction Dataset from Kaggle Repository | Accuracy 96% by RF |
| Harshita et al. [22] | LR, SVM, RF, K-NN and Decision Tree | Stroke Prediction Dataset from Kaggle Repository | Accuracy 95.5% by RF and LR |
| Dev et al. [23] | Neural Network (NN), RF and Decision Tree | Stroke Dataset from Electronic Medical Record Repository | Accuracy 77% by NN |
| Shoily et al. [24] | K-NN, J48, NB and Random Forest | Self-made dataset of 1058 patients | Accuracy 85.6% by NB |
| Li et al. [25] | LR, NB, DT, Bayesian Network and Random Forest | China National Stroke Screening Data (2017) | Precision 97% by RF and Recall 99.94% by DT |
| Govindarajan et al. [26] | Artificial Neural Network (ANN), SVM, LR, DT, Bagging and Boosting | Stoke Data from Sugam Multispecialty Hospital, India | Accuracy-95.3% by ANN |
| Hager et al. [27] | LR, RF, DT and Support Vector Machine | Healthcare Stroke Data from Kaggle Repository | Accuracy 90% by RF |
| Emon et al. [28] | LR, DT, K-NN, Weighted Voting, MLP, XGBoost, AdaBoost and Gradient Boosting | Healthcare Stroke Data from Kaggle Repository | Accuracy 97% by Weighted Voting |
| Mustafa et al. [30] | RF, K-NN, NB, SVM, LR and Stacking | Healthcare Stroke Data from Kaggle Repository | Accuracy 97% by Stacking |
| Kumar [31] | XGBoost | Healthcare Stroke Data from Kaggle Repository | Accuracy 98% |
| Gupta et al. [32] | LR, NB, K-NN, DT, Adaboost, RF and XGBoost | Healthcare Stroke Data from Kaggle Repository | Accuracy 97.22% by RF |
| Zubaidai et al. [33] | RF, DT, LR, SVM, Hard Voting and Soft Voting | Healthcare Stroke Data from Kaggle Repository | Accuracy 94.6% by RF |

TABLE II. RESULTS' COMPARISON OF THE DIFFERENT STUDIES ON THE HEALTHCARE STROKE DATASET FROM KAGGLE REPOSITORY

| Study | Algorithm | Performance Metric | Accuracy | Recall | Precision | F-measure |
|---|---|---|---|---|---|---|
| Sailasya et al. [19] | Naive Bayes | Accuracy | 82% | 85.7% | 79.2% | 82.3% |
| Sharma et al. [20 | Random Forest | Accuracy | 98.94% | 98.9% | 99% | 99% |
| Tazin et al. [21] | Random Forest | Accuracy | 96% | 94.9% | 97% | 95.93% |
| Hager et al. [27] | Random Forest | Accuracy | 90% | 90.5% | 91.5% | 91% |
| Emon et al. [28] | Weighted Voting | Accuracy | 97% | 96.12% | 99.4% | 97.73% |
| Mostafa et al. [30] | Stacking | Accuracy | 97% | 97.37% | 96.93% | 97.14% |
| Kumar. [31] | XGBOOST | Accuracy | 98% | 94.6% | 97.6% | 96.07% |
| Gupta et al. [32] | Random Forest | Accuracy | 97% | 94.9% | 99.6% | 97% |
| Zubaidai et al. [33] | Random Forest | Accuracy | 94.7% | 95.94% | 93.16% | 94.52% |
| Dritsas et al. [37] | Stacking | Accuracy | 98% | 97.4% | 97.4% | 97.4% |

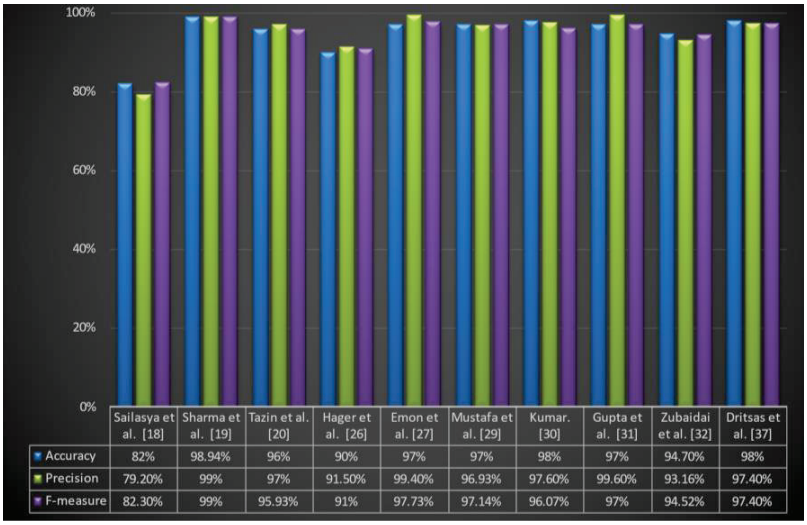



Fig. 2. Comparison of different studies on Healthcare Kaggle stroke dataset

The analysis of the studies also revealed several important factors that may influence the effectiveness of predictive models for identifying individuals at risk of experiencing a stroke. For example, models that included more predictors tended to be more accurate than those with fewer predictors.

Additionally, models that were based on the machine learning techniques, such as random forest or Boosting technique, tended to perform better than those based on traditional statistical methods. While the findings of this review are generally consistent with previous reviews and meta-analyses, several limitations should be noted. First, the quality of studies varied, with some studies having a high risk of bias. Second, studies included in this review focused primarily on short-term outcomes, and there is a need for more research on the long-term effectiveness of predictive models for stroke prevention. Finally, there is a need for more research on the optimal combination of predictors and the best methods for integrating these predictors into clinical practice.

## V. Research Gaps

- The dataset used in earlier research were imbalanced, which prevents the model from correctly identifying the minority class without affecting the accuracy of the model, so we need to use techniques that can handle this type of imbalanced dataset [20, 22, 24, 25, 27, 28].
- The attributes that were considered till now on this topic were quite few. In order to improve the model's performance, other medical attributes such as systolic, diastolic, and pulse pressure may be taken into account [16, 20, 24, 25] [27-32].
- In the final implementation of the stroke dataset, [16, 21, 22] have considered only accuracy as the performance metrics for the model evaluation and have not considered other performance metrics like Recall, Precision, ROC Curve.
- None of the researcher in their research work considered the Model execution time as the parameter for the model evaluation [16-32].

## VI. Conclusion

Early detection of stroke illnesses can help high-risk persons make lifestyle changes, decreasing complications and perhaps saves lives, which might be a huge medical advancement. Stroke is a very dangerous disease that requires immediate cure to avoid further complications. In this review paper we have provided a literature review of several stroke detection techniques. Different machine learning approaches were used and we found that Random Forest, Logistic Regression and Neural Network have the best accuracy and prediction. On the basis of this study, we will try to formulate a machine learning model for early identification of Brain Stroke aiming at high accuracy at a low cost and complexity.

## References


[1] "About stroke," Centers for Disease Control and Prevention, 02-Nov-2022. [Online]. Available: https://www.cdc.gov/stroke/about.htm.[Accessed: 16-Nov-2022].

[2] "Stroke," Mayo Clinic, 20-Jan-2022. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/stroke/symptoms-causes/syc-20350113. [Accessed: 09-Jan-2023].

[3] World Health Organization. [Online]. Available: http://www.emro.who.int/health-topics/stroke-cerebrovascular-accident/index.html. [Accessed: 16-Nov-2022].

[4] "Risk factors for stroke," Risk Factors for Stroke — Johns Hopkins Medicine, 13-Dec-2022. [Online]. Available: https://www.hopkinsmedicine.org/health/conditions-and-diseases/stroke/risk-factors-for-stroke. [Accessed: 08-Jan-2023].

[5] "Ischemic stroke," Medline Plus. [Online]. Available: https://www.medlineplus.gov/ischemicstroke.html. [Accessed: 17-Nov-2022].

[6] "The top 10 causes of death," World Health Organization, 09-Dec-2020. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. [Accessed: 08-Feb-2023].

[7] "Hemorrhagic strokes (bleeds)," [Online]. Available: https://www.stroke.org/en/about-stroke/types-of-stroke/hemorrhagic-strokes-bleeds. [Accessed: 17-Nov-2022].

[8] Y. Kumar, A. Koul, R. Singla, and M. F. Ijaz, "Artificial Intelligence in disease diagnosis: A systematic literature review, synthesizing framework and future research agenda," Journal of Ambient Intelligence and Humanized Computing, 2022.

[9] S. Das, A. Dey, A. Pal, and N. Roy, "Applications of artificial intelligence in machine learning: Review and Prospect," International Journal of Computer Applications, vol. 115, no. 9, pp. 31–41, Apr 2015.

[10] D. Petersson, "What is supervised learning?" Enterprise AI, 26-Mar-2021. [Online]. Available: https://www.techtarget.com/searchenterpriseai/definition/supervised- learning. [Accessed: 12-Jan-2023].

[11] R. Choudhary and H. K. Gianey, "Comprehensive Review on Supervised Machine Learning Algorithms," 2017 International Conference on Machine Learning and Data Science (MLDS), pp. 37–43, Dec 2017.

[12] K. K. Verma, B. M. Singh, and A. Dixit, "A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in Intelligent surveillance system," International Journal of Information Technology, vol. 14, no. 1, pp. 397–410, 2019.

[13] V. Rajyaguru, C. Vithalani, and R. Thanki, "A literature review: Various learning techniques and its applications for eye disease identification using retinal images," International Journal of Information Technology, vol. 14, no. 2, pp. 713–724, 2020.

[14] K. K. Verma, B. M. Singh, and A. Dixit, "A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in Intelligent surveillance system," International Journal of Information Technology, vol. 14, no. 1, pp. 397–410, 2019.

[15] M. A. Ramessur and S. D. Nagowah, "A predictive model to estimate effort in a sprint using machine learning techniques," International Journal of Information Technology, vol. 13, no. 3, pp. 1101–1110, 2021.

[16] J. N. Heo, J. G. Yoon, H. Park, Y. D. Kim, H. S. Nam, and J. H. Heo, "Machine learning–based model for prediction of outcomes in acute stroke," Stroke, vol. 50, no. 5, pp. 1263–1265, May 2019.

[17] R. S. Jeena and S. Kumar, "Stroke prediction using SVM," 2016 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), pp. 600–602, Dec. 2016.

[18] F. S. Alotaibi, "Implementation of machine learning model to predict heart failure disease," International Journal of Advanced Computer Science and Applications, vol. 10, no. 6, pp. 261–268, 2019.

[19] G. Sailasya and G. L. Kumari, "Analyzing the performance of stroke prediction using ML classification algorithms," International Journal of Advanced Computer Science and Applications, vol. 12, no. 6, pp. 539–545, 2021.

[20] C. Sharma, S. Sharma, M. Kumar, and A. Sodhi, "Early stroke prediction using machine learning," 2022 International Conference on Decision Aid Sciences and Applications (DASA), pp. 890–894, Mar. 2022.

[21] T. Tazin, M. N. Alam, N. N. Dola, M. S. Bari, S. Bourouis, and M. Monirujjaman Khan, "Stroke disease detection and prediction using robust learning approaches," Journal of Healthcare Engineering, vol. 2021, pp. 1–12, Nov. 2021.

[22] H. K V, H. P, G. Gupta, V. P, and P. K B, "Stroke prediction using machine learning algorithms," International Journal of Innovative Research in Engineering & Management, vol. 8, no. 4, pp. 6–9, 2021.

[23] S. Dev, H. Wang, C. S. Nwosu, N. Jain, B. Veeravalli, and D. John, "A predictive analytics approach for stroke prediction using machine learning and Neural Networks," Healthcare Analytics, vol. 2, pp. 1–9, Nov. 2022.

[24] T. I. Shoily, T. Islam, S. Jannat, S. A. Tanna, T. M. Alif, and R. R. Ema, "Detection of stroke disease using machine learning algorithms," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1–6, Jul. 2019.

[25] X. Li, D. Bian, J. Yu, M. Li, and D. Zhao, "Using machine learning models to improve stroke risk level classification methods of China National Stroke Screening," BMC Medical Informatics and Decision Making, vol. 19, no. 1, pp. 1–7, Dec. 2019.

[26] P. Govindarajan, R. K. Soundarapandian, A. H. Gandomi, R. Patan, P. Jayaraman, and R. Manikandan, "Classification of stroke disease using machine learning algorithms," Neural Computing and Applications, vol. 32, no. 3, pp. 817–828, Feb. 2020.

[27] H. Ahmed, E. M. G. Younis, A. A. Ali, and S. ghany, "Stroke Prediction using Distributed Machine Learning Based on Apache Spark," International Journal of Advanced Science and Technology, vol. 28, no. 15, pp. 89–97, 2019.

[28] M. U. Emon, M. S. Keya, T. I. Meghla, M. M. Rahman, M. S. Mamun, and M. S. Kaiser, "Performance Analysis of Machine Learning Approaches in stroke prediction," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 1464–1469, Nov. 2020.

[29] M. Kaur, S. R. Sakhare, K. Wanjale, and F. Akter, "Early stroke prediction methods for prevention of Strokes," Behavioural Neurology, vol. 2022, pp. 1–9, Apr. 2022.

[30] S. A. Mostafa, D. S. Elzanfaly, and A. E. Yakoub, "A machine learning ensemble classifier for prediction of Brain Strokes," International Journal of Advanced Computer Science and Applications, vol. 13, no. 12, pp. 258–266, 2022.

[31] A. K. Uttam, "Analysis of Uneven Stroke Prediction dataset using machine learning," 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1209–1213, May 2022.

[32] S. Gupta and S. Raheja, "Stroke prediction using machine learning methods," 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 553–558, Jan. 2022.

[33] H. Al-Zubaidi, M. Dweik, and A. Al-Mousa, "Stroke prediction using machine learning classification methods," 2022 International Arab Conference on Information Technology (ACIT), pp. 1–8, Nov. 2022.

[34] O. Almadani and R. Alshammari, "Prediction of stroke using data mining classification techniques," International Journal of Advanced Computer Science and Applications, vol. 9, no. 1, pp. 457–460, 2018.

[35] M. Monteiro, A. C. Fonseca, A. T. Freitas, T. Pinho e Melo, A. P. Francisco, J. M. Ferro, and A. L. Oliveira, "Using machine learning to improve the prediction of functional outcome in ischemic stroke patients," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 15, no. 6, pp. 1953–1959, Mar. 2018.

[36] S. Rahman, M. Hasan, and A. K. Sarkar, "Prediction of brain stroke using machine learning algorithms and deep neural network techniques," European Journal of Electrical Engineering and Computer Science, vol. 7, no. 1, pp. 23–30, Jan. 2023.

[37] E. Dritsas and M. Trigka, "Stroke risk prediction with machine learning techniques," Sensors, vol. 22, no. 13, pp. 1–13, Jun. 2022.