



Early Prediction of Stroke Based on Deep and Machine Learning by Applying Medical Imaging and Tabular Data

Qabeela Q. Thabit^{1*}, Alyaa I. Dawood², Bayadir A. Issa²

¹ Basrah Education Directorate, Ministry of Education, Basrah 61015, Iraq

² Engineering Technical College-Basrah, Southern Technical University, Basrah 61018, Iraq

Corresponding Author Email: qabelh2010@gmail.com

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/mmep.111228>

ABSTRACT

Received: 3 February 2024

Revised: 20 June 2024

Accepted: 5 July 2024

Available online: 31 December 2024

Keywords:

stroke, machine learning methods, artificial neural networks, deep learning, Jupyter Notebook, Python language, different dataset, prediction

Stroke is a serious disease that causes a high mortality rate annually. People who have previously had a stroke are exposed to blindness, paralysis, and confusion, as well as an increased rate of recurrence. Every year 15 million people in the world are infected, 5 million of whom die, and 5 million others remain permanently disabled, which makes this disease a danger to society if ways are developed to reduce it. Therefore, work to reduce the infection has drawn the attention of many researchers. This disease is diagnosed in the early stages before a stroke occurs. This work provides early diagnosis before stroke using various artificial intelligence algorithms, while diversifying the use of disease-specific databases in more than one way to detect and determine the best methods to suggest it. In this work, we presented 8 algorithms for machine learning, deep learning using three- and four-layer neural networks, and finally a convolutional neural network using the VGG 16 model. The methods mentioned were applied to various datasets, including Kaggle (The health care problem: Predicting stroke patients) and MRI images. Machine learning algorithms of up to 100% and deep learning of up to 99% in medical imaging have made early recognition of stroke possible and essential for diagnosing and understanding this deadly disease.

1. INTRODUCTION

In the past 20 years, stroke has become one of the top causes of mortality and lifelong disability worldwide. Only in China, there are 2 million patients diagnosed with stroke annually, and the mortality rate is 11.48%. Stroke is divided into ischemic kinds, which account for 85% of cases, and hemorrhagic types, which account for 15% of cases. Patients who have had a stroke are subject to long-term moral and financial responsibilities if they recover from the condition while others pass away as a result of a stroke surgery [1].

Stroke victims have a better chance of recovering from strokes if they seek medical attention as soon as possible. Any stroke victim should visit a doctor as soon as possible. Otherwise, death, long-term disability, and brain damage will occur. Various factors can cause strokes in patients. A stroke is caused primarily by diet, inactivity, alcohol, tobacco, personal history, medical history, and complications, according to the National Heart, Lung, and Blood Institute [2].

Due to the development of the field of artificial intelligence with its multiple branches of machine learning and deep learning, which includes neural, convoluted, and recurrent networks, as well as the field of data science and its analysis, in anticipating many health problems and detecting diseases in the early stages of the development of these diseases, stroke was and still is the subject of interest and research by many researchers Those who focused their efforts on this disease and

the early detection of it because of its unfortunate consequences for people and society [3, 4].

Artificial intelligence research needs data analysis tools in performing tasks for large data, i.e., the sample approved by the responsible authorities is of a large size, and therefore, recently, solid sites have begun to publish data from approved and licensed agencies by the World Health Organization for research purposes, and the data volume may reach up to millions of people, and this is what makes the need for advanced and efficient algorithms with a high capacity to process all this data [5, 6].

This data includes a lot of information about patients, including personal information such as name, identification number, gender, and others related to the clinical symptoms shown in patients, while some of them include percentages of blood analyses, thyroid, associated infections, or medical history, in addition to other research that uses radiological images, scans, or magnetic resonance [7, 8].

Stroke is one of the diseases that has occurred recently. It poses a great danger due to the number of deaths that the disease causes in society, in addition to the fact that the person being treated is exposed to symptoms that accompany it, which may last a lifetime, such as loss of speech, paralysis, or other damages. The case of sudden cessation of blood flow constitutes a burden on people. Medical personnel are a case without introductions, so artificial intelligence models have been designed to predict stroke. In our proposed work, we

emphasize early detection of stroke, which contributes to reducing the harm of this deadly disease, according to the World Health Organization [1, 2].

This paper focuses on predicting stroke with machine learning algorithms as well as with deep learning and comparing which of the results is better. Then, apply the principle of basic data analysis and enter it for processing again in those algorithms, and also compare the difference in the results obtained by the two methods, as this is a contribution to the field of practical application. For those algorithms, moreover, in both machine and deep learning algorithms, it will make changes to the parameters used, for example, the number of hidden layers and the type of optimization function.

In this work, we provided early diagnosis of stroke based on data, whether it was in the form of an Excel sheet in which several data were collected for a group of people, including one who was affected and the other was healthy, or it was magnetic resonance images, which contributes to the possibility of reaching good results by diversifying methods and databases. To change the parameters of each model to achieve the desired results. This study presents a model for diagnosing stroke with Our experiment showed good results, and it is possible to use these models in the future in health institutional contexts and integrate them with the mechanisms used in hospitals to produce rapid and accurate diagnostic results. An interactive programming environment Anaconda was used with a graphical interface and a textual interface supported by python-newt, which is freely available on the Internet and can also be easily downloaded to a personal laptop. Jupyter Notebook application: It is easy to use and follow the steps. It is worth noting that Jupyter Notebook is a programming application in the Python language.

The rest of the paper is organized as follows: The second section includes the works related to this paper; the third section describes the data used; the fourth section includes the presented methodology; the fifth section includes prediction algorithms; the sixth contains measures of achievement and results; the seventh compares previous works and the current work; and finally, conclusions and recommendations for future work are presented in the eighth section of this paper.

2. RELATED WORKS

The optimal performance of the proposed models requires large amounts of data, and this requires high computing capabilities, including graphics and data processing. In addition, the problem of transparency still reflects a challenge and a dilemma in interpreting the results obtained. One of the first researches that dealt with the issue of stroke prediction using machine learning was by Liu et al. [9], and after that, research continued in this field, where Rahman et al. [5] presented a set of machine learning algorithms, and the best of them in performance was the Random Forest, where they achieved a result of 98%. Soumyabrata Dev et al. [10] also presented a model for stroke prediction based on machine learning with neural networks supported by data pre-processing through basic data analysis.

Fang et al. [1] used the freely available International Stroke Trial (IST) dataset to apply three contemporary deep learning (DL) algorithms to predict the 6-month Ischemic Stroke (IS) outcome. The performance of these machine learning algorithms and machine learning (ML) for clinical prediction

was another objective of this study. The comparison between different ML techniques (Deep Forest, Random Forest, Support Vector Machine, etc.) and the DL frameworks that are already in use (CNN, LSTM, Resnet) reveals that DL is not noticeably superior to ML. To evaluate organized medical data, Desert Locust methods and reporting need to be established and enhanced.

Other studies have used brain scans from Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) and MRI scans that display various infarct types to identify a stroke. The method's outcomes were assessed visually. That demonstrated the potential for identifying strokes and showed that MRI is superior to CT in this regard [11]. While other researchers resorted to merging the two types of algorithms, machine and deep learning, and comparing their results to get the best performance [12, 13].

3. DATASET TYPES DESCRIPTION

In the work presented here, we used two types of data for training in order to benefit from the different types of data and how to reach results through these different databases.

The first was an Excel sheet that was downloaded from the well-known website "Kaggle: HealthCare Problem: Prediction Stroke Patients". The sheet included 5110 rows and 12 columns, which were deleted. Of these, the first column is not useful in diagnosis, which is the patient's identification number. There are 10 columns remaining that represent the characteristics, and the eleventh column is the last column, which represents the output and includes whether there is a disease or not.

The dataset is explained in an abstract manner, and the contribution to the proposed work represents the use of more than one database dataset 1 that employed two types of algorithms (machine learning algorithms and artificial neural networks in different number layers). The first is the medical records of the patients, and the second is a set of magnetic resonance images.

The medical records included in the paper presented are the dataset in experiment one, which is related to machine learning and deep learning, "Kaggle: HealthCare Problem: Prediction Stroke Patients," which is actually the patients' medical records, such as age, gender, hypertension, heart disease, and physiological and environmental information as shown in Table 1 and Table 2.

There are 5110 rows in dataset 1, each row representing a patient, 12 columns containing 10 features, an identification (ID) column, and a target feature column with either a border (1) or no border (0). With 4861 normal patients and 249 stroke patients in the dataset, it will be processed later during the training phase to find the balance [14]. Information about this dataset is explained in Tables 1 and 2 [14].

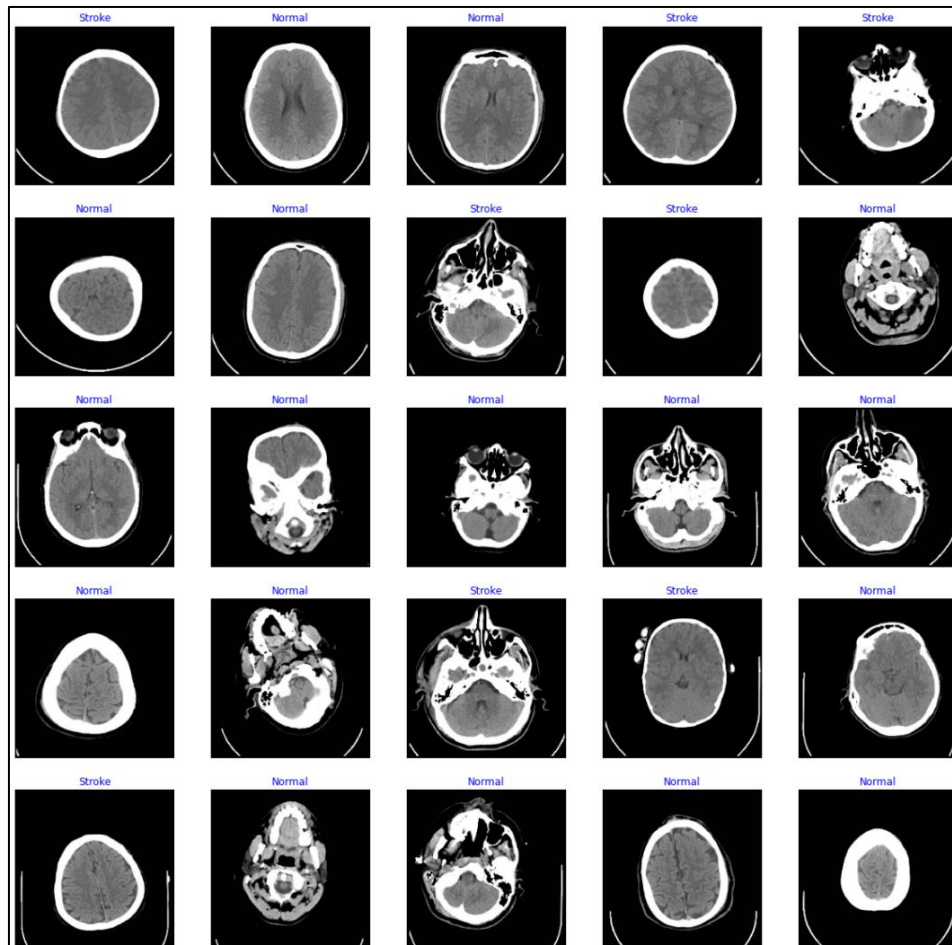
The second experiment involves deep learning in a convolutional neural network. The dataset used in this paper is MRI images, which include images of two types of people: normal and stroke brain cases [15]. The data set contains 2515 images, divided into 1551 normal images and 964 images of cerebral hemorrhage. The images were recorded in a resolution format of up to 256×256 pixels for each image. Figure 1 describes a different image for normal cases and injured cases, which were also obtained from the Kaggle website [15].

Table 1. Dataset 1 information used in proposed work

Name of Column	Description
ID Individuals	Unique for each one of 5110
Gender	Female = 1 while Male = 0
Age	Patient Age (1-82)
Hypertension	If patient has hypertension (1) or not (0)
Heart Disease	Indicating whether the unique patient's id has a heart problem (1) or not (0)
Ever Married	Displays the marital status as either yes (1) or no (0)
Work Type	It displays the patient's five different work status categories. If Children = 0, Job Government = 1, No worked = 2, Private = 3, Self-Worked = 4
Residence Type	Whether Rural = "0" or Urban = "1," it defines the sort of living area.
Average Glucose Level	Average Glucose Range (55.12-271.74)
BMI	Body Mass Index Range (10.3-97.6)
Smoking Status	There are four different types of smokers listed. Smokes = "2"; Never Smoked = "1"; Previously Smoked = "0"; "3" for unknown (Related to the unknown type, no information is available or could be located.)
Stroke	Stroke shows if the target has a stroke (1) or not (0).

Table 2. Dataset information

Index	Age	Hypertension	Heart Disease	Avg Glucose Level	BMI	Stroke	Gender	Ever_Married	Work Type	Residence Type	Smoking Status
0	67.0	0	1	228.69	36.6	1	1	1	2	1	1
1	61.0	0	0	202.21	28.1	1	0	1	3	0	2
2	80.0	0	1	105.92	32.5	1	1	1	2	0	2
3	49.0	0	0	171.23	34.4	1	0	1	2	1	3
4	79.0	1	0	174.12	24.0	1	0	1	3	0	2
5	81.0	0	0	186.21	29.0	1	1	1	2	1	1
6	74.0	1	1	70.09	27.4	1	1	1	2	0	2
7	69.0	0	0	94.39	22.8	1	0	0	2	1	2
8	59.0	0	0	76.15	28.1	1	0	1	2	0	0
9	78.0	0	0	58.57	24.2	1	0	1	2	1	0

**Figure 1.** Dataset 2 (normal and stroke brain) images

4. PROPOSED METHODOLOGY

4.1 Data preprocessing

There are several steps that must be taken, which are preparation or initialization of the database. It is necessary to investigate first if there are missing values or outliers, and this represents an obstacle to processing the data if it has been processed from the start.

The characteristics that affect the output must be investigated. For example, some of the features in the table of the database do not have any effect, such as the identification number ID of the patient. This is neglected or removed from the data, and this stage is called the selection of features. The next stage is the stage of scaling the features, in which it can be reduce the differences between the numbers that represent these features within a specific standard, then the data is divided into data for training and others for testing, and this is what is known as data splitting. This indicates that the initial stage of the work is to adjust the data used in order to train the model more effectively and efficiently.

Cleaning the data is one of the most important steps after collecting the data, which are divided into twelve characteristics. The ID column is ignored because it does not affect the construction of the model. One of the important steps that must be taken is to check the data if there are missing or empty values, and the BMI column contains empty cells. In this case, the most common value can be placed, or an arithmetic average of the values can be taken and placed in the blanks, also from the important steps. Here, the literal data is converted to numeric to be understood by the computer. In this type of database, there is data that must be converted to digital, such as gender, married or not, type of work, and smoking status. If this data is converted to digital, it is difficult to understand and receive it by the algorithms used.

The data set approved in this work consists of 5110 rows, and of these rows, 4861 of them are for cases that do not suffer from stroke, while 249 are aunts, and this represents an imbalance in the data. When using this data, it may obtain high accuracy, but other accuracy measures such as retrieval and accuracy may not be sufficient, and therefore the results will be incorrect and the expectations obtained are important. The dataset was balanced using the Synthetic Minority Over-sampling Technique (SMOTE) method for an effective model by using the Random Oversampling (ROS) approach, which is one of the Python tools to balance the data.

To limit the features between -1 and 1 in order to normalize them, MinMaxScaler tools were used. Next, the principle of statistical principal component analysis (PCA) was used, which chooses the minimum number of principal components. After the unbalanced data set has been managed and the data has been prepared, the model construction step starts. 20% of the data are test data, while the remaining 80% are training data. The accuracy and effectiveness of this function are improved by training the model with a variety of post-split classification methods.

In several applications of machine learning algorithms, such as deep learning with a neural network in several layers and performing training with a convolutional neural network, the results are compared in all cases in order to come up with the best results and Figure 2 explains the procedure for working.

5. PREDICTION ALGORITHMS

In fact, we present in the paper the algorithms that resulted in the best results in order not to explain all the algorithms without meaning to their results. Therefore, the presentation here is of the best among all machine learning algorithms.

As for neural networks, we still obtained good results through three layers and quarter layers, so there is no need to waste time and effort by implementing neural networks with more layers. As for the deep learning network, which is the convolutional neural network, specifically the VGG 16 convolutional neural network, which is distinguished in terms of performance because it has depth. Almost as good, as it only has 16 depth layers that have weights instead of relying on a large number of redundant parameters. Moreover, the selection of parameters for training of the compiler contains specifications of: optimizer and loss.

We find that it showed the best results in the other path because the training was done by adjusting the weights, and thus we obtained a high accuracy rate and a very low data loss rate to reach the highest level of control without deficiency or over-fitting.

5.1 Machine learning methods

There are various methods for detecting disease cases, whether they affect humans, plants, or animals, but machine learning methods remain the main field of research in this topic, as they have established the foundations for these studies [16, 17]. The following are the machine learning algorithms that were implemented in this paper:

- (1) K-Nearest Neighbors (KNN): K-NN is a form of slow learning where no pre-processing stage is specified and all calculations are saved for classification. This data categorization technique makes decisions based on the nearby training data points on the feature map. The target

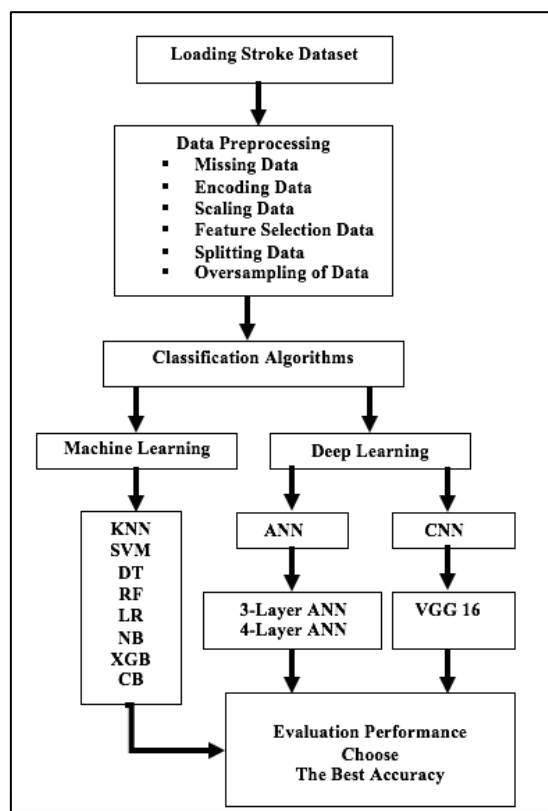


Figure 2. Procedure of working

class is predicted by the K-NN classifier using the Euclidean distance metric. The best value for parameter k, which governs how well the classifier performs, is determined by the dataset. After examining the impacts, the optimal value is then established [18].

- (2) Support Vector Machine (SVM): The support vector machine is one of the supervised machine learning techniques that categorize unknown input using labeled data. Decision limits are expressed using the notion of decision levels, or hyperplanes. It is important to note that a set of data objects can be divided into many classifications using the superlevel. The supported vector machine was employed in this study to categorize the data by developing a function that, with the least amount of error and maximum percentage of accuracy, labels each data point with the proper label [19, 20].
- (3) Decision Tree (DT): Since this study focuses on stroke prediction, decision trees can be utilized as both classification and regression methods until all features, including data, are finished, each stage of the contract is concluded with a question-and-answer session [21, 22].
- (4) Random Forest (RF): The so-called random selection forest, sometimes known as the random forest, is a classification and regression procedure that, in reality, consists of a lot of decision trees. Each tree offers a solution to a particular issue. The best choice is to classify each tree according to its output, which is the best solution from the set of tree solutions. In order to improve the performance of classifiers and the accuracy of classification or prediction, group learning provides a more accurate method. In order to achieve this, a random forest classifier is used, in which many decision trees are grouped on different subsets of data and averaged to improve the accuracy expected [4, 5].
- (5) Logistic Regression (LR): The logistic regression technique is used in classification problems, where the output is in the form of probability, which is then translated into one of the categories. Upon hearing about this algorithm, the initial assumption is that it is utilized for solving regression issues. It is a method that uses the sigmoid function to estimate the relationship between pairs of dependent variables as well as at least one individual variable, and it is amenable to supervised training. It is used to resolve regression issues in addition to classification issues. It might fall under one of the three categories in terms of classification [7, 19].
 - Binary classification
 - Multiple classifications
 - Ordinal classification
- (6) Naïve Bayes (NB): One of the methods of supervised machine learning, which relies in its work on the principle of Bayes' theory, is the probability theory that predicts the occurrence of an event based on a prior set of conditions associated with that event [20, 21].
- (7) XGBoost: XGboost is a fantastic example of a gradient augmentation technique. The gradient gain alternative can be carefully planned for accuracy and optimization, even though there may not be any ground-breaking mathematical breakthroughs in this scenario. It consists of a linear version, and the newborn tree may be a technique that uses various AI algorithms to determine whether a flimsy newborn would produce a reliable newborn in order to improve the correctness of the version. For

instance, parallel learning and impulsive learning can both be used to learn random forest [19].

- (8) CatBoost: Hancock and Khoshgoftaar's [23] gradient boosting technique known as CatBoost works well with categorical characteristics by reducing information loss. CatBoost is distinct from other gradient-boosting techniques. It initially uses ordered boosting, a useful variant of gradient boosting methods, to handle the problem of target leakage. This technique also does well with small datasets. Thirdly, CatBoost can control category features. This handling is typically completed in the preprocessing stage and comprises converting the original category variables to one or more numerical values [23, 24].

5.2 Deep learning methods

Deep learning is learning by means of artificial neural networks when it includes several layers called by this name, and deep learning is learning using neural networks and it has three types according to the type of network and how to process it [25-27].

- Deep learning using neural networks
- Deep learning using convolutional networks
- Deep learning using iterative networks

Deep neural networks with several layers are used in order to obtain better results and high accuracy. As the number of layers of the neural network increases, the complexity increases, and the possibility of obtaining high accuracy in training and examining data increases [28, 29]. The artificial neural network consists of several layers starting from the input layer and the middle layers known as the hidden layers and the output layer, the big change is usually made in the middle layers where more than one layer is used and the parameters used are changed in it from the number of neurons and the function used, as well as what is known as cutting neural connections with the function known as dropout. This paper presents deep learning for both types of data used for such purposes. Once the data is trained in the form of Excel sheets, and another time the data is trained in the form of magnetic resonance images (MRI) by using convolutional neural networks (CNN), especially VGG16, as shown in Figure 3, then the results obtained from all methods were compared [30].

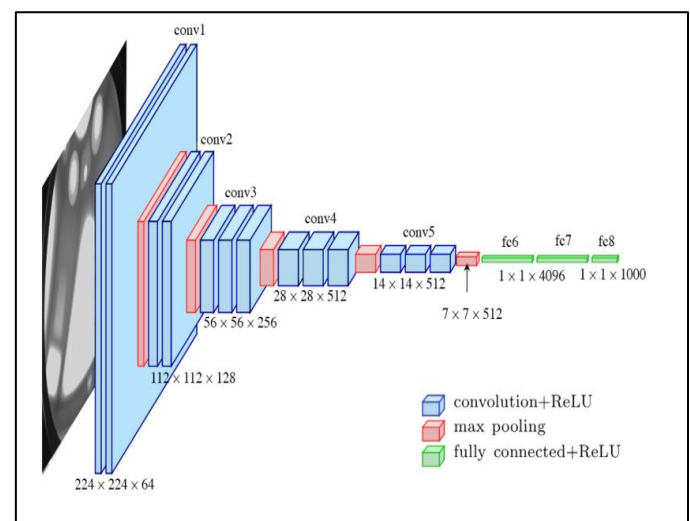


Figure 3. VGG-16 components

6. MEASUREMENTS OF ACHIEVEMENT AND RESULTS

6.1 Metrics evaluation

This work used five statistical variables—accuracy, precision, recall/sensitivity (recall and sensitivity in binary classification are the same), F1 Score, and RUC AUC Score to assess the effectiveness and utility of the classifiers [5, 19, 31, 32]. The statistical parameter definitions are as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{F1 Score} = \frac{2 * (\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})} \quad (3)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (4)$$

TN: method predicted stroke as true stroke.

TP: method predicted normal as(normal), true predicted.

FN: method predicted the normal as(stroke), false predicted.

FP: method predicted stroke as(normal), false predicted.

6.2 Obtained results

Experiment 1: Include two parts: The dataset is trained using machine learning algorithms and obtained the results mentioned in Tables 3-5. The algorithm SVM showed the highest accuracy.

Experiment 2: Include MRI scan image dataset.

Figures 4-8 compare all the results we obtained from all the models completed in this paper, starting with the ratio Precision, Recall, F1-Score, RUC AUC Score, Accuracy, respectively.

Table 3. Obtained results from machine learning algorithms

Algorithm	Precision	Recall	F1-Score	RUC AUC Score	Accuracy
KNN	0.98	0.98	0.98	0.97694	0.97685
SVM	1.0	1.0	1.0	1.0	1.0
DT	0.98	0.98	0.98	0.9769	0.9768
RF	0.9999	0.9999	0.9999	0.9999	0.99485
LR	0.97	0.96	0.97	0.97	0.97
NB	0.98	0.98	0.98	0.98	0.98
XGB	0.98	0.98	0.98	0.99839	0.9778
CatB	0.97	0.96	0.96	0.9972	0.9645

Table 4. Obtained results from deep learning algorithms (ANN)

Algorithm	Precision	Recall	F1-Score	RUC AUC Score	Accuracy
3-layer	0.96	0.95	0.96	0.9512	0.95632
4-layer	0.97	0.96	0.96	0.9661	0.9649

Table 5. Obtained results from VGG-16

Algorithm	Precision	Recall	F1-Score	RUC AUC Score	Accuracy
VGG 16-4 layer	0.9999	0.9999	0.9999	0.9999	0.9999

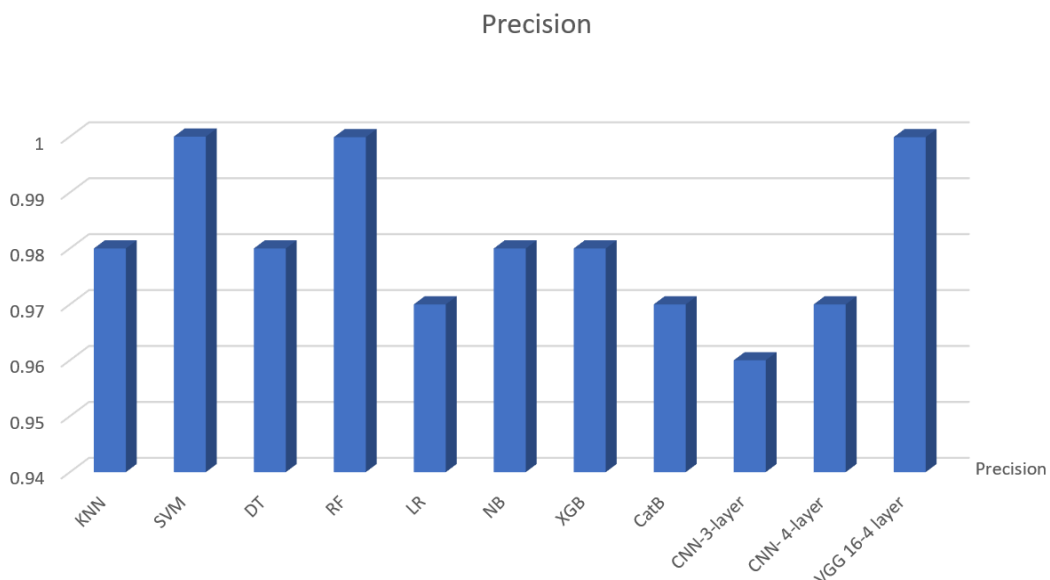


Figure 4. Precision comparison

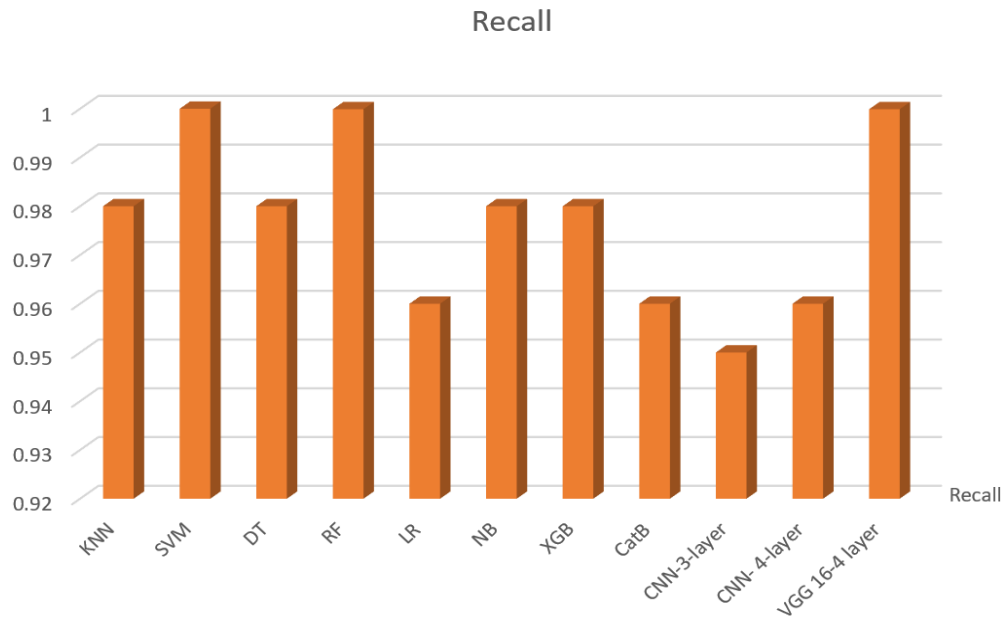


Figure 5. Recall comparison

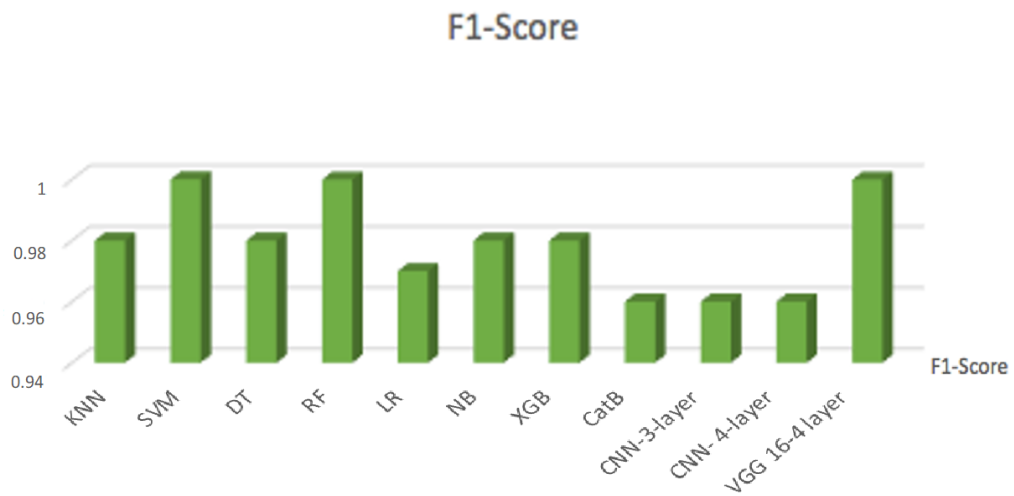


Figure 6. F1-Score comparison

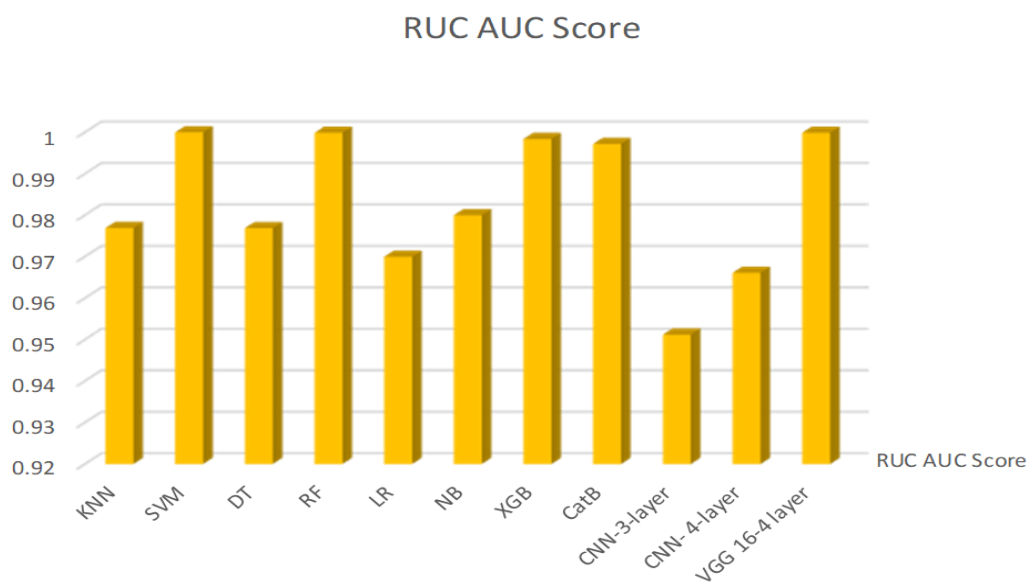


Figure 7. RUC AUC Score comparison

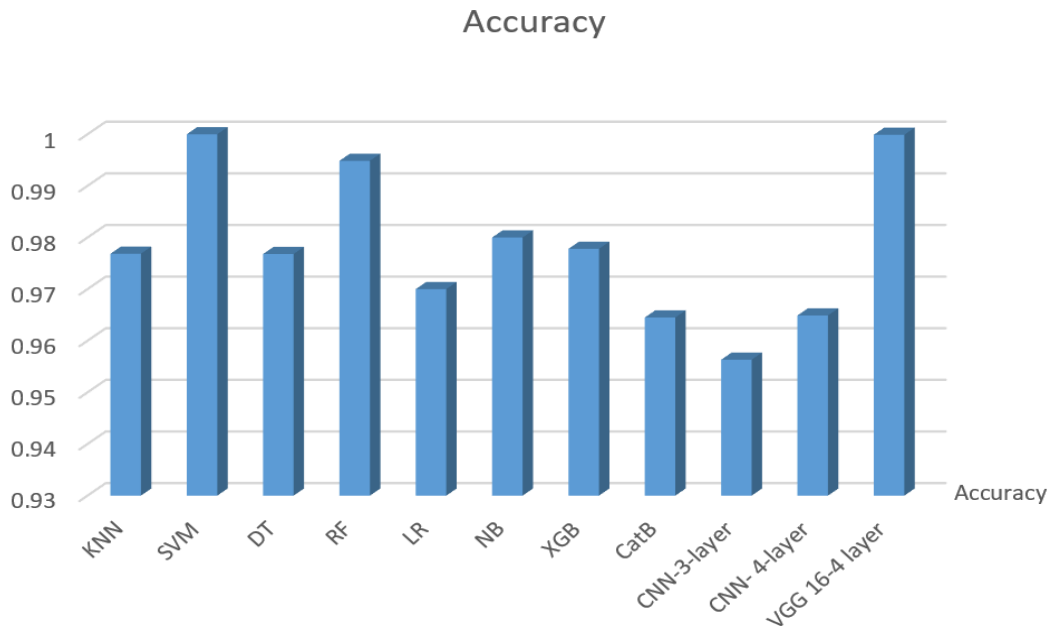


Figure 8. Accuracy comparison

7. COMPARISON WITH PREVIOUS WORKS

machine learning and deep learning were compared to previous results, as shown in the Table 6.

The results we obtained from applying algorithms for

Table 6. Comparison results with previous works

Ref.	Data Source Type	Execution Algorithm	Accuracy
[1]	The International Stroke Trial (IST) dataset	LSTM and Resnet (CNN)	0.83
[2]		biLSTM and LSTM	91%
[3]	Biopac Systems Inc., Goleta, CA, USA	Adaptive Gradient Boosting, XGBoost, LightGBM	87%
			0.80%
			0.77%
[5]	Kaggle: HealthCare Problem: Prediction Stroke Patients	LR, DT, RF, KNN, SVM, GaussianNB, BernoulliNB, XGBoost, AdaBoost, LGBM	0.78%
			0.71%, 0.98%, 0.99%, 0.96%, 0.82%, 0.70%, 0.67%, 0.97%, 0.78%, 0.95%
			0.871%, 0.868%, 0.866%, 0.839%
[13]	National Health Insurance Research Database (NHIRD)	DNN, GBDT, LR, SVM	0.95%, 0.947%, 0.926%, 0.875%, 0.95%, 0.955%
[18]	National Health Insurance Research Database (NHIRD)	LR, RF, DT, NB, SVM, CNN	0.968%
[27]	The MRI dataset	Deep learning (AlexNet)	0.9113%
[33]	Second Affiliated Hospital of Fujian Medical University	VGG-16	0.84%
[34]	The MRI dataset	Bidirectional LSTM-bidirectional LSTM-ATT	1.0%
Proposed work	Kaggle: HealthCare Problem: Prediction Stroke Patients	Highest among machine learning (SVM)	0.9649%
		Deep Neural Network	0.9999%
		CNN-VGG-16	0.9999%

8. CONCLUSIONS

Given the seriousness of stroke, it is one of the diseases that cause the death of many people, sometimes without the appearance of previous symptoms of the disease, so it is important to study the possibility of helping medical tools commonly used in treatment to predict the disease and support them with the features that are available with machine and deep learning algorithms because of their potential in Diagnosing and predicting results by feeding them with a data set after the model is well trained. Thus, artificial intelligence

algorithms have shown amazing results in this aspect. In addition, resorting to using different types of databases leads to the creation of different works, and there is an idea to expand the database with data processing methods in the near future by merging more than one database, and it is possible to transfer the outputs of machine learning results as input to deep learning models, and in The paper presented here showed that the results obtained reached 96%, 99%, and 100% accuracy through deep learning using neural networks, convolutional network VGG-16, and machine learning using several algorithms, respectively. One of the works that can be

applied to what is presented here is to combine machine and deep learning, thus combining the different data used, and the results can also be expanded by applying a performance measurement model for all results combined. Important suggestions in this area that can be added to future recommendations are as follows:

- Conduct ablation studies to evaluate the contribution of different components.
 - Validation of models in other datasets for stroke prediction.
- Investigating model group and transfer learning approaches.

REFERENCES

- [1] Fang, G., Huang, Z., Wang, Z. (2022). Predicting ischemic stroke outcome using deep learning approaches. *Frontiers in Genetics*, 12: 827522. <https://doi.org/10.3389/fgene.2021.827522>
- [2] Kaur, M., Sakhare, S.R., Wanjale, K., Akter, F. (2022). Early stroke prediction methods for prevention of strokes. *Behavioural Neurology*, 2022(1): 7725597. <https://doi.org/10.1155/2022/7725597>
- [3] Islam, M.S., Hussain, I., Rahman, M.M., Park, S.J., Hossain, M.A. (2022). Explainable artificial intelligence model for stroke prediction using EEG signal. *Sensors*, 22(24): 9859. <https://doi.org/10.3390/s22249859>
- [4] Sirsat, M.S., Fermé, E., Câmara, J. (2020). Machine learning for brain stroke: A review. *Journal of Stroke and Cerebrovascular Diseases*, 29(10): 105162. <https://doi.org/10.1016/j.jstrokecerebrovasdis.2020.105162>
- [5] Rahman, S., Hasan, M., Sarkar, A.K. (2023). Prediction of brain stroke using machine learning algorithms and deep neural network techniques. *European Journal of Electrical Engineering and Computer Science*, 7(1): 23-30. <https://doi.org/10.24018/ejece.2023.7.1.483>
- [6] Owaid, M.A., Hammoodi, A.S. (2024). Evaluating machine learning and deep learning models for enhanced DDoS attack detection. *Mathematical Modelling of Engineering Problems*, 11(2): 493-499. <https://doi.org/10.18280/mmep.110221>
- [7] Emon, M.U., Keya, M.S., Meghla, T.I., Rahman, M.M., Al Mamun, M.S., Kaiser, M.S. (2020). Performance analysis of machine learning approaches in stroke prediction. In *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, pp. 1464-1469. <https://doi.org/10.1109/ICECA49313.2020.9297525>
- [8] Magris, M., Iosifidis, A. (2023). Bayesian learning for neural networks: An algorithmic survey. *Artificial Intelligence Review*, 56(10): 11773-11823. <https://doi.org/10.1007/s10462-023-10443-1>
- [9] Liu, T., Fan, W., Wu, C. (2019). A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset. *Artificial Intelligence in Medicine*, 101: 101723. <https://doi.org/10.1016/j.artmed.2019.101723>
- [10] Dev, S., Wang, H., Nwosu, C.S., Jain, N., Veeravalli, B., John, D. (2022). A predictive analytics approach for stroke prediction using machine learning and neural networks. *Healthcare Analytics*, 2: 100032. <https://doi.org/10.1016/j.health.2022.100032>
- [11] Dey, N., Rajinikanth, V. (2022). Automated detection of ischemic stroke with brain MRI using machine learning and deep learning features. In *Magnetic Resonance Imaging, Recording, Reconstruction and Assessment Primers in Biomedical Imaging Devices and Systems*, pp. 147-174. <https://doi.org/10.1016/b978-0-12-823401-3.00004-3>
- [12] Hossain, D., Scott, S.H., Cluff, T., Dukelow, S.P. (2023). The use of machine learning and deep learning techniques to assess proprioceptive impairments of the upper limb after stroke. *Journal of NeuroEngineering and Rehabilitation*, 20(1): 15. <https://doi.org/10.1186/s12984-023-01140-9>
- [13] Ganakwar, P., Date, M.S. (2020). Convolutional neural network-VGG16 for road extraction from remotely sensed images. *International Journal for Research in Applied Science and Engineering Technology*, 8(8): 916-922. <https://doi.org/10.22214/ijraset.2020.30796>
- [14] Stroke Prediction Dataset. <https://www.kaggle.com/fedesoriano/stroke-prediction-dataset>.
- [15] Brain Stroke Prediction CT Scan Image Dataset. <https://www.kaggle.com/datasets/iashiqu/brain-stroke-prediction-ct-scan-image-dataset>.
- [16] Albadri, R.F., Awad, S.M., Hameed, A.S., Mandeel, T.H., Jabbar, R.A. (2024). A diabetes prediction model using hybrid machine learning algorithm. *Mathematical Modelling of Engineering Problems*, 11(8): 2119-2126. <https://doi.org/10.18280/mmep.110813>
- [17] Arianti, N.D., Muslih, M., Irawan, C., Saputra, E., Sariyusda, Bulan, R. (2023). Classification of harvesting age of mango based on NIR spectra using machine learning algorithms. *Mathematical Modelling of Engineering Problems*, 10(1): 204-211. <https://doi.org/10.18280/mmep.100123>
- [18] Ashrafuzzaman, M., Saha, S.K., Nur, K. (2022). Prediction of stroke disease using deep CNN based approach. *Journal of Advances in Information Technology*, 13(6). <https://doi.org/10.12720/jait.13.6.604-613>
- [19] Letham, B., Rudin, C., McCormick, T.H., Madigan, D. (2015). Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model. *The Annals of Applied Statistics*, 9(3): 1350-1371. <https://doi.org/10.1214/15-aos848>
- [20] Thabit, Q.Q. (2023). Deep and machine learning for improving breast cancer detection. *Engineering and Technology Journal*, 8(12): 3156-3163. <https://doi.org/10.47191/etj/v8i12.06>
- [21] Kadali, D.K., Mohan, R.N.V.J., Naik, M.C., Bokka, Y. (2024). Crime data analysis using Naive Bayes classification and least square estimation with MapReduce. *International Journal of Computational Methods and Experimental Measurements*, 12(3): 289-295. <https://doi.org/10.18280/ijcmem.120309>
- [22] Dritsas, E., Trigka, M. (2022). Stroke risk prediction with machine learning techniques. *Sensors*, 22(13): 4670. <https://doi.org/10.3390/s22134670>
- [23] Hancock, J.T., Khoshgoftaar, T.M. (2020). CatBoost for big data: An interdisciplinary review. *Journal of Big Data*, 7(1): 94. <https://doi.org/10.1186/s40537-020-00369-8>
- [24] García, L., Tomás, J., Parra, L., Lloret, J. (2019). An m-health application for cerebral stroke detection and monitoring using cloud services. *International Journal of Information Management*, 45: 319-327.

- <https://doi.org/10.1016/j.ijinfomgt.2018.06.004>
- [25] Su, S.S., Li, L.Y., Wang, Y., Li, Y.Z. (2023). Stroke risk prediction by color Doppler ultrasound of carotid artery-based deep learning using Inception V3 and VGG-16. *Frontiers in Neurology*, 14: 1111906. <https://doi.org/10.3389/fneur.2023.1111906>
- [26] Guo, Z., Li, X., Huang, H., Guo, N., Li, Q. (2019). Deep learning-based image segmentation on multimodal medical imaging. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 3(2): 162-169. <https://doi.org/10.1109/TRPMS.2018.2890359>
- [27] Iqbal, S., Qureshi, A.N., Li, J., Mahmood, T. (2023). On the analyses of medical images using traditional machine learning techniques and convolutional neural networks. *Archives of Computational Methods in Engineering*, 30(5): 3173-3233. <https://doi.org/10.1007/s11831-023-09899-9>
- [28] Thabit, Q.Q., Dawood, A.I., Issa, B.A. (2021). Implementation three-step algorithm based on signed digit number system by using neural network. *Indonesian Journal of Electrical Engineering and Computer Science*, 24(3): 1832-1839. <http://doi.org/10.11591/ijeecs.v24.i3.pp1832-1839>
- [29] Rajendran, K., Radhakrishnan, M., Viswanathan, S. (2022). An ensemble deep learning network in classifying the early CT slices of Ischemic Stroke patients. *Traitement du Signal*, 39(4): 1089-1098. <https://doi.org/10.18280/ts.390401>
- [30] Al-Mekhlafi, Z.G., Senan, E.M., Rassem, T.H., Mohammed, B.A., Makbol, N.M., Alanazi, A.A., Almurayziq, T.S., Ghaleb, F.A. (2022). Deep learning and machine learning for early detection of stroke and haemorrhage. *Computers, Materials and Continua*, 72(1): 775-796. <https://doi.org/10.32604/cmc.2022.024492>
- [31] Chantamit-O-Pas, P., Goyal, M. (2017). Prediction of stroke using deep learning model. In *Neural Information Processing: 24th International Conference, ICONIP, Guangzhou, China, pp. 774-781*. https://doi.org/10.1007/978-3-319-70139-4_78
- [32] Thabit, Q.Q., Fahad, T.O., Dawood, A.I. (2022). Detecting diabetes using machine learning algorithms. In *2022 Iraqi International Conference on Communication and Information Technologies (IICCIT), Basrah, Iraq, pp. 131-136*. <https://doi.org/10.1109/IICCIT55816.2022.10010408>
- [33] Choi, Y.A., Park, S.J., Jun, J.A., Pyo, C.S., Cho, K.H., Lee, H.S., Yu, J.H. (2021). Deep learning-based stroke disease prediction system using real-time bio signals. *Sensors*, 21(13): 4269. <https://doi.org/10.3390/s21134269>
- [34] Zhou, Y.D., Gong, Z., Li, L. (2023). Deep learning-based multi-feature auxiliary diagnosis method for early detection of ischemic stroke. *Traitement du Signal*, 40(2): 433-443. <https://doi.org/10.18280/ts.400203>