Fourth International Conference on Electronics, Communication and Aerospace Technology (ICECA-2020)
IEEE Xplore Part Number: CFP20J88-ART; ISBN: 978-1-7281-6387-1

Performance Analysis of Machine Learning Approaches in Stroke Prediction

Minhaz Uddin Emon*, Maria Sultana Keya†, Tamara Islam Meghla‡, Md. Mahfujur Rahman§,
M Shamim Al Mamun¶, and M Shamim Kaiser‖
*†§Dept of Computer Science and Engineering
Daffodil International University, Dhaka-1207, Bangladesh
Email:*minhazkhondokar21@gmail.com, †maria.sultana.keya@gmail.com, §mrrajuiit@gmail.com
‡Software, Web & Cloud, Faculty of Information Technology and Communications, Tampere University, Finland
Email:‡tamara.meghla@tuni.fi
¶‖Institute of Information Technology, Jahangirnagar University, Dhaka, Bangladesh
Email:¶shamim@juniv.edu, ‖mskaiser@juniv.edu

***Abstract*—Most of strokes will occur due to an unexpected obstruction of courses by prompting both the brain and heart. Early awareness for different warning signs of stroke can minimize the stroke. This research work proposes an early prediction of stroke diseases by using different machine learning approaches with the occurrence of hypertension, body mass index level, heart disease, average glucose level, smoking status, previous stroke and age. Using these high features attributes, ten different classifiers have been trained, they are Logistics Regression, Stochastic Gradient Descent, Decision Tree Classifier, AdaBoost Classifier, Gaussian Classifier, Quadratic Discriminant Analysis, Multi layer Perceptron Classifier, KNeighbors Classifier, Gradient Boosting Classifier, and XGBoost Classifier for predicting the stroke. Afterwards, results of the base classifiers are aggregated by using the weighted voting approach to reach highest accuracy. Moreover, the propsoed study has achieved an accuracy of 97%, where the weighted voting classifier performs better than the base classifiers. This model gives the best accuracy for the stroke prediction. The area under curve value of weighted voting classifier is also high. False positive rate and false negative rate of weighted classifier is lowest compared with others. As a result, weighted voting is almost the perfect classifier for predicting the stroke that can be used by physicians and patients to prescribe and early detect a potential stroke.**

***Keywords*—Stroke, Machine Learning, Confusion Matrices, Area Under Curve (AUC), Weighted Voting, Correlation Matrix**

I. Introduction

A stroke will occur when the blood flow to various areas of the brain is disrupted or diminished, the cells in those regions do not get the nutrients and oxygen and start to die. A stroke is a medical emergency which requires immediate care. Early detection and proper management is required to minimize the further damage in the affected area of the brain and other complication in the body parts. According to World Health Organization (WHO) in every year fifteen million people are suffering from stroke in worldwide and affected individuals are passing away every 4-5 minutes.

The two forms of strokes are ishemic and hemorrhagic. In the event of an ischemical stroke, drainage is blocked by clots, and in the event of a hemorrhagic stroke, a weak blood vessel explodes and bleeds into the brain. Stroke can be prevented by a healthy/balanced lifestyle that is wiping off the bad lifestyle like smoking and drinking, controlling body mass index (BMI) and average glucose level, maintaining good health of heart and kidney. The prediction of stroke is necessary and shall be treated to prevent permanent damage or death. This paper has considered hypertension, BMI level, heart disease, and average glucose level as parameters for predicting stroke. In addition, machine learning can play a vital role in the decision making processes of the proposed prediction system [1]–[3].

In the literature, very few recorded research works have used machine learning models to predict stroke [4]–[9]. The machine learning algorithms are artificial neural network (ANN), stochastic gradient descent, c4.5 decision tree algorithm, k-nearest neighbor (kNN), principle component analysis (PCA), convolutional neural network (CNN), naive bayes etc. A relation is correlated among the diseases/attributes such as hypertension, BMI level, average glucose level, and heart disease with stroke [10].

Our contribution in this paper is as follows-

- A weighted voting classifier is proposed in predicting stroke using the diseases/attributes such as hypertension, body mass index level, heart disease, average glucose level, smoking status, previous stroke and age.
- A performance of the proposed weighted voting classifier is compared with the state-of-the-art classifier such as Logistics Regression (LR), Stochastic Gradient Descent (SGD), Decision Tree Classifier (DTC), AdaBoost, Gaussian, Quadratic Discriminant Analysis (QDC), Multi Layer Perceptron (MLP), KNeighbors, Gradient Boosting Classifier (GBC), XGBoost (XGB).

The rest of the paper is organized as following. Section 2 discusses some literature review on the existing research. Research methodologies are stated in section 3 and it is separated as three parts: data description, machine learning classifiers and evaluation matrices, implementation procedures are discussed. In section 4, result and discussion are shown and the details will describe about the correlation result and performance analysis. Finally, the conclusion is discussed in section 5.

II. Literature Review

Many researchers have already used machine learning based approached to predict strokes. Govindarajan et al. [11] conducted a study to categorize stroke disorder using a text mining combination and a machine learning classifier and collected data for 507 patients. For their analysis, they used various machine learning approaches for training purposes using ANN, and the SGD algorithm gave them the best value, which was 95%.

Amini et al. [4], [12] conducted research to predict stroke incidence, collected 807 healthy and unhealthy subjects in their study categorized 50 risk factors for stroke, diabetes, cardiovascular disease, smoking, hyperlipidemia, and alcohol use. They used two techniques that had the best accuracy from c4.5 decision tree algorithm, and it was 95%, and for K-nearest neighbor, the accuracy was 94%.

Cheng et al. [13] published a report on the estimation of the ischemic stroke prognosis. In their analysis, 82 ischemic stroke patient data were used, two ANN models were used to find precision, and 79% and 95% were used.

Cheon et al. [14]–[16] performed a study to predict stroke patient mortality. In their study, they used 15099 patients to identify stroke occurrence. They used a deep neural network approach to detect strokes. The authors used PCA to extract medical record history and predict stroke. They have got an area under the curve (AUC) value of 83%.

Singh et al. [17] performed a study on stroke prediction applied to artificial intelligence. In their research, they used a different method for predicting stroke on the cardiovascular health study (CHS) dataset. And they took the decision tree algorithm to feature extract to principal component analysis. They used a neural network classification algorithm to construct the model they got 97% accuracy.

Chin et al. [18] performed a study to detect an automated early ischemic stroke. In their study, the main purpose was to develop a system using CNN to automated primary ischemic stroke. They collected 256 images to train and test the CNN model. In their system image prepossessing remove the impossible area that can't occur of stroke, they used the data prolongation method to raise the collected image. Their CNN method has given 90% accuracy. Sung et al. [5] performed a study to develop a stroke severity index. They collected 3577 patient's data with acute ischemic stroke. For their predicting models, they used various data mining techniques and linear regression. Their prediction feature got the best result from the k-nearest neighbor model (95% CI).

Monteiro et al. [19] performed a study to get a functional outcome prediction of ischemic stroke using machine learning. In their research, they apply this technique to a patient who was passing three months after admission. They got the AUC value above 90%.

Kansadub et al. [20] performed a study to predict stroke risk. In the study, the authors employed Naive Bayes, Decision Tree, and Neural Network to analyze data to predict stroke. In their study, they used accuracy and AUC as their pointer's assessment. All of this algorithm, they classified decision tree and naive Bayes gave the most accurate.

Adam et al. [21] performed a study to classify ischemic stroke. They used two models: a k-nearest neighbor and a decision tree algorithm to classified ischemic stroke. In their research, the decision tree algorithm was more usable for medical specialists who used it to classify stroke.

III. Research Methodology

This section is divided into three parts, these are: Data description, machine learning classifiers & evaluation matrices, implementation procedures. These three processes are described below:

A. Data Description

In this paper, the informational collection utilized has been acquired from the medical clinic of Bangladesh. It's the document of 5110 people's information and now all the attributes are described:

age: This attribute means a person's age. It's numerical data.

gender: This attribute means a person's gender. It's categorical data.

hypertension: This attribute means that this person is hypertensive or not. It's numerical data.

work_type: This attribute represents the person work scenario. It's categorical data.

residence_type: This attribute represents the person living scenario. It's categorical data.

heart_disease: This attribute means whether this person has a heart disease person or not. It's numerical data.

avg_glucose_level: This attribute means what was the level of a person's glucose condition. It's numerical data.

bmi: This attribute means body mass index of a person. It's numerical data.

ever_married: This attribute represents a person's married status. It's categorical data.

smoking_Status: This attribute means a person's smoking condition. It's categorical data.

stroke: This attribute means a person previously had a stroke or not. It's numerical data.

In this all attribute stroke is the decision class and rest of the attribute is response class.

B. Machine Learning Classifiers & Evaluation Matrices

This section discusses ten machine learning classifiers, which are used here to build stroke predictors. And this classifiers list are: (1)LR, (2)SGD, (3)DTC, (4)AdaBoost, (5)Gaussian, (6)QDA, (7)MLP, (8)KNeighbors, (9)GBC, (10)XGB . The reason behind choosing these classifiers is that these are well known classifiers in building vulnerability predictors and used in several similar research work. These ten classifiers are selected for building vulnerability predictors in our model, this well known classifiers are used several research work [22], [23], as similar of ours. Moreover, these models are evaluated by measuring the confusion matrices.

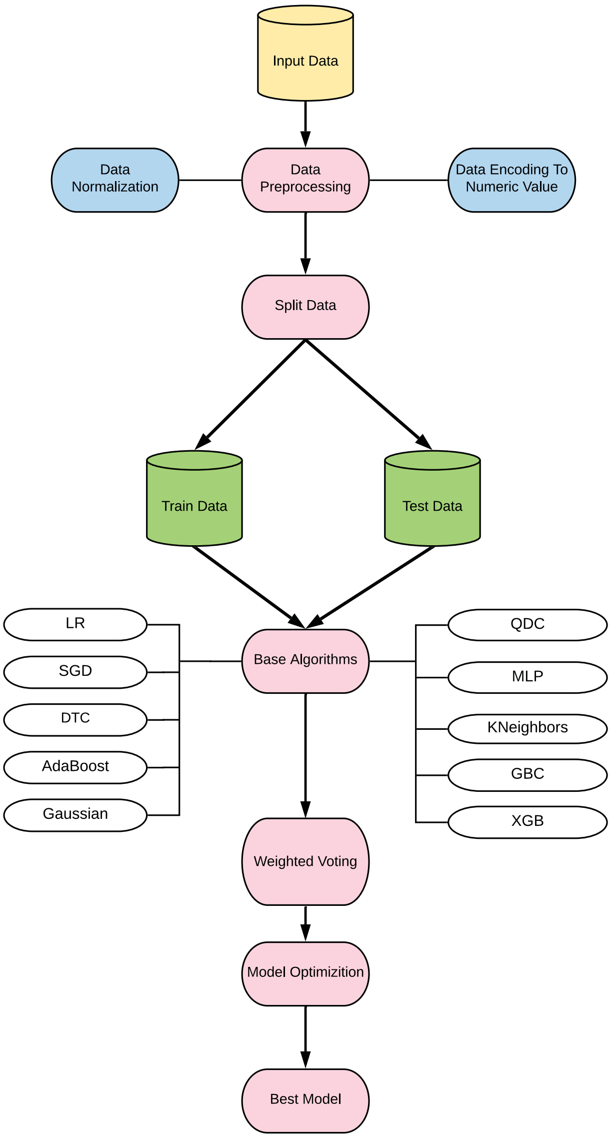


Fig. 1. Procedure of stroke prediction.

Fig. 1 mention the graphical representation of procedure for predicting stroke of different algorithms in step by step.

C. Implementation Procedure:

The implementation procedure is described in this section. To complete the study, Python and Scikit-learn libraries have been utilized and all the procedure presented in figure 1.

1) **Input Data:** The 5110 patient details are collected based on their various health conditions, which is in the occurrence of stroke disease. The data is collected from many hospitals of Bangladesh. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The study proceeded after receiving approval from the noninvasive ethical committee of the Jahangirnagar University (JU), Dhaka, Bangladesh. All participants were presented with the necessary documentation required to comply with the ethical standard approved by the JU.
2) **Data prepossessing:** For processing data, firstly it checks the missing value and duplicate value. Missing values were filled up by taking the mean/Median of the other values. Some missing value is detected in smoking status attribute. These missing value fill up with group by age attribute. In our dataset, there is no duplicate value. Then it normalized our data set and label encoding to the categorical data. Then it will find all the data set as numerical value. Finally, the standard data set is obtained for further processing.
3) **Split Data:** Splitting a dataset means separating this data into two types: training and testing. In this paper,split technique is used for training & testing purpose.
4) **Base Algorithms:** Ten algorithms are taken as a base algorithm to train and test proposed approach.
5) **Weighted Voting**: After implementing all the classifier, weighted voting classifier is implemented to improve the accuracy of all the classifier's.
6) **Model Optimization:** In this procedure, confusion matrix is measured of each model to find out the value of precision, recall, f-1 score, auc, FP Rate and FN Rate.
7) **Best Model:** In this procedure, the level of accuracy of ten algorithm is measured of dataset to produce accuracy of different types of algorithms and find the best model using weighted voting classifier.

IV. Results & Discussion

A. Correlation Results

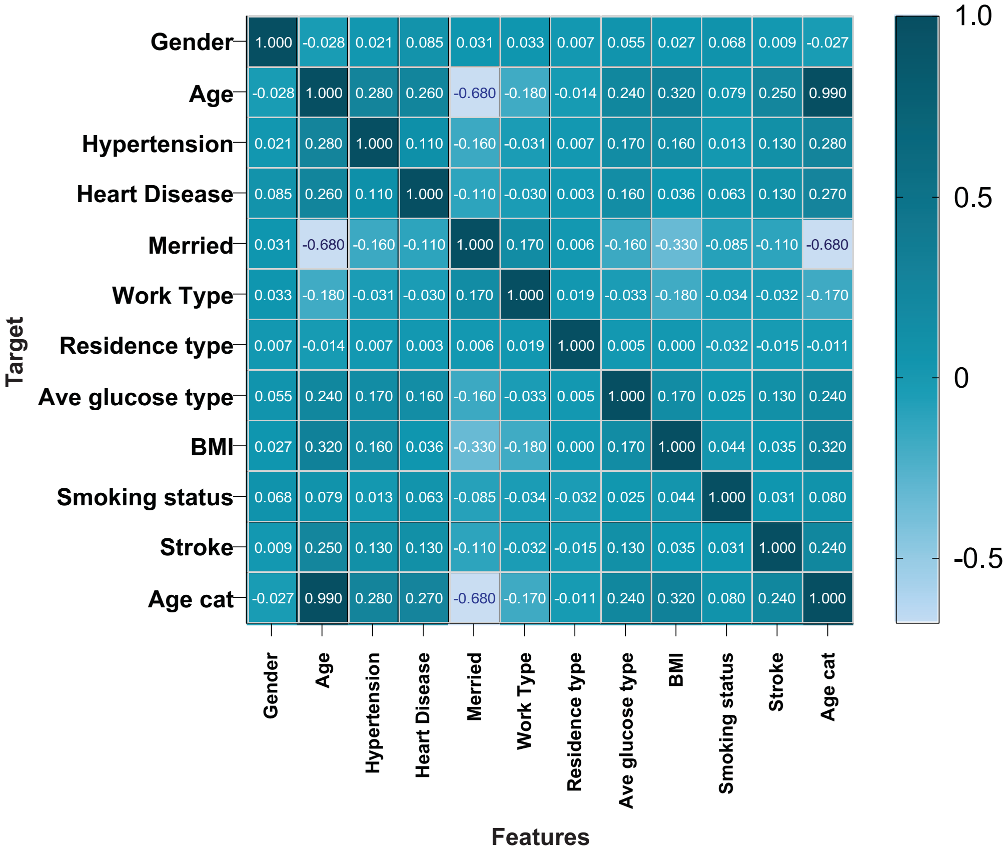
The consequences of Pearson connection uncovers the effect of feature attributes on target attribute. Figure 2 visualizes the connection between stroke attribute and others attribute. From the figure, obviously no single metric profoundly effect on stroke. Among the metrics gender, age, hypertension, heart_disease, avg_glucose_level, bmi, smoking_status have respectably high effect on stroke. The least effect factors are ever_married, work_type, residence_type.

B. Performance Analysis

The results section will discuss the test dataset, which is used for machine learning approach and find out their accuracy to classify the data. Among 5110 data, 1556 data has been used as testing purposes.

Table I represents confusion matrices of the stroke prediction using ten different classifiers, namely: LR, SGD, DTC, AdaBoost, Gaussian, QDA, MLP, KNeighbors, GBC, XGB and weighted voting classifier for measuring the performance of stroke prediction.

In Table II, this paper apply ten classifiers to predict stroke performance, after applying ten classifier, weighted voting



	Gender	Age	Hypertension	Heart Disease	Merried	Work Type	Residence type	Ave glucose type	BMI	Smoking status	Stroke	Age cat
Gender	1.000	-0.028	0.021	0.085	0.031	0.033	0.007	0.055	0.027	0.068	0.009	-0.027
Age	-0.028	1.000	0.280	0.260	-0.680	-0.180	-0.014	0.240	0.320	0.079	0.250	0.990
Hypertension	0.021	0.280	1.000	0.110	-0.160	-0.031	0.007	0.170	0.160	0.013	0.130	0.280
Heart Disease	0.085	0.260	0.110	1.000	-0.110	-0.030	0.003	0.160	0.036	0.063	0.130	0.270
Merried	0.031	-0.680	-0.160	-0.110	1.000	0.170	0.006	-0.160	-0.330	-0.085	-0.110	-0.680
Work Type	0.033	-0.180	-0.031	-0.030	0.170	1.000	0.019	-0.033	-0.180	-0.034	-0.032	-0.170
Residence type	0.007	-0.014	0.007	0.003	0.006	0.019	1.000	0.005	0.000	-0.032	-0.015	-0.011
Ave glucose type	0.055	0.240	0.170	0.160	-0.160	-0.033	0.005	1.000	0.170	0.025	0.130	0.240
BMI	0.027	0.320	0.160	0.036	-0.330	-0.180	0.000	0.170	1.000	0.044	0.035	0.320
Smoking status	0.068	0.079	0.013	0.063	-0.085	-0.034	-0.032	0.025	0.044	1.000	0.031	0.080
Stroke	0.009	0.250	0.130	0.130	-0.110	-0.032	-0.015	0.130	0.035	0.031	1.000	0.240
Age cat	-0.027	0.990	0.280	0.270	-0.680	-0.170	-0.011	0.240	0.320	0.080	0.240	1.000

Fig. 2. Correlation matrices among the Socio-demographics, lifestyle status and disease.

TABLE I
CONFUSION MATRICES FOR MACHINE LEARNING CLASSIFIERS TO PREDICT STROKE

Classifiers Name	Predicted→	No Stroke	Stroke
	Actual↓		
LR	No stroke	811	176
	Stroke	165	404
SDG	No Stroke	478	509
	Stroke	28	541
DTC	No Stroke	914	73
	Stroke	56	513
AdaBoost	No Stroke	963	24
	Stroke	59	510
Gaussian	No Stroke	756	231
	Stroke	105	464
QDA	No Stroke	770	217
	Stroke	104	465
MLP	No Stroke	744	243
	Stroke	78	491
KNeighbors	No Stroke	803	184
	Stroke	17	552
GBC	No Stroke	986	1
	Stroke	48	514
XGB	No Stroke	984	3
	Stroke	45	524
Weighted Voting	No Stroke	992	5
	Stroke	40	529

classifier is used to evaluate the results and measure the accuracy value, precision value, recall value, f-1 score and auc value, which is highest from other classifier and this is 97%, the second highest accuracy is obtained from GBC, XGB classifier is 96% and third highest accuracy is obtained from AdaBoost classifier and value is 94%. The lowest accuracy is obtained from SGD classifier and this is 65%.

Moreover, this paper analyses the another metrics of performance analysis, which are false positive rate and false negative rate. The higher FN rate indicates false alarms which are created by the model. In terms of the false positive rate and the false negative rate, the figure 3 indicates the relative success of classifiers. So the best predictor among selected techniques is weighted voting classifier.

On the other hand, The Area Under Curve (AUC) is the indicator of a classification's ability to discriminate between classes. If AUC = 1, so all positives and negatives class points can be properly distinguished by the classifier. However, if the AUC is 0, both negative and positive would be expected by the classifier as positive.There is a high probability of the classifier being able to differentiate positive class values from negative class values when $0.5 < AUC < 1$ is used. This

TABLE II
MEASUREMENT RESULT FOR ML CLASSIFIER TO PREDICTING STROKE

CN	Accuracy	Class Label	Precision	Recall	F-1	AUC	FP Rate	FN Rate
LR	78%	**No Stroke**	0.83	0.82	0.83	0.76	30%	16%
		Stroke	0.70	0.71	0.71			
SGD	65%	**No Stroke**	0.68	0.95	0.80	0.73	48%	5%
		Stroke	0.76	0.26	0.39			
DTC	91%	**No Stroke**	0.94	0.92	0.93	0.80	12%	5%
		Stroke	0.87	0.90	0.89			
AdaBoost	94%	bNo Stroke	0.92	0.98	0.95	0.79	4%	5%
		Stroke	0.97	0.85	0.91			
Gaussian	78%	**No Stroke**	0.86	0.77	0.81	0.77	33%	12%
		Stroke	0.97	0.78	0.72			
QDA	79%	**No Stroke**	0.87	0.79	0.73	0.75	31%	11%
		Stroke	0.69	0.80	0.84			
MLP	79%	**No Stroke**	0.91	0.79	0.85	0.81	33%	9%
		Stroke	0.71	0.88	0.78			
KNeighbors	87%	**No Stroke**	0.97	0.83	0.89	0.81	25%	2%
		Stroke	0.77	0.96	0.95			
GBC	96%	**No Stroke**	0.93	0.99	0.96	0.85	0.1%	4%
		Stroke	0.99	0.87	0.93			
XGB	96%	**No Stroke**	0.94	0.99	0.97	0.90	0.5%	4%
		Stroke	0.99	0.89	0.94			
Weighted Voting	97%	**No Stroke**	0.93	1.00	0.97	0.93	0.9%	3%
		Stroke	1.00	0.90	0.95			

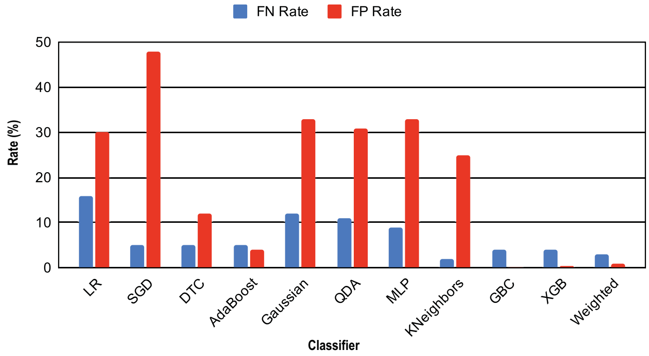


Fig. 3. FP & FN rate of different Classifiers.

is because both True Positive and True Negative statistics are found than False negative or False positives. The classifier is unable to differentiate between positive and negative class points when AUC = 0.5 is used. The classifier either estimates a random class or a constant class over all data points. The AUC for LR, SGD, DTC, AdaBoost, Gaussian, QDA, MLP, KNN, GBC, XGB are 0.76, 0.73, 0.80, 0.79, 0.77, 0.75, 0.81, 0.81, 0.85, 0.90, 0.93 respectively and Weighted Voting AUC value is 0.93.

Table III shows that, there are lot of existing approaches is used to predict stroke by ML classifiers and also deep learning. So, here some state-of-art methods and their accuracy are compared with our proposed model and it is noticed that the proposed study has achieved an accuracy of 97%.

TABLE III
PERFORMANCE COMPARISON OF STROKE PREDICTION MODEL

Ref	Method Name	Accuracy
Govindarajan et al [11]	NLP-ML	95%
Amini et al [4]	C4.5 DT, KNN	95%, 94%
Cheng et al [13]	ANN	79%, 95%
Cheon et al [14]	DNN	83%
Singh et al [17]	ANN	96%
Chin et al [18]	CNN	90%
Sung et al [5]	KNN	95%
Proposed Method	Proposed Weighted Voting	97%

Legend: NLP– Natural Language Processing; DNN– Deep Neural Network

V. CONCLUSION

The proposed research work has employed ten classifiers to find out the performance of stroke occurrence in a person. The proposed weighted voting classifier has considered gender, age, hypertension, heart disease, average glucose level, BMI, smoking status feature attributes to predict stroke. The performance evaluation reveals that weighted voting provided the highest accuracy of about 97% compared to the commonly used other machine learning algorithms. As a result, the weighted voting can be considered for the prediction of stroke. The relationship between these diseases and possibility of occurring stroke in a human individual has been evaluated. So, if this disease is diagnosed and maintained correctly from early stage, then it will help to reduce the occurrence of stroke in our life. In the future, deep learning based imaging, such as brain CT scan and MRI, can be proposed together with an existing model to boost the performance indices.

REFERENCES

[1] M. Mahmud *et al.*, “A brain-inspired trust management model to assure security in a cloud based iot framework for neuroscience applications,” *Cognitive Computation*, vol. 10, no. 5, pp. 864–873, 2018.

[2] M. B. T. Noor, N. Z. Zenia, M. S. Kaiser, S. Al Mamun, and M. Mahmud, “Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection

of alzheimer's disease, parkinson's disease and schizophrenia," *Brain Informatics*, vol. 7, no. 1, pp. 1–21, 2020.

[3] M. Mahmud, M. S. Kaiser, and A. Hussain, "Deep learning in mining biological data," *arXiv preprint arXiv:2003.00108*, 2020.

[4] L. Amini, R. Azarpazhouh, M. T. Farzadfar, S. A. Mousavi, F. Jazaieri, F. Khorvash, R. Norouzi, and N. Toghianfar, "Prediction and control of stroke by data mining," *International Journal of Preventive Medicine*, vol. 4, no. Suppl 2, pp. S245–249, May 2013.

[5] S.-F. Sung, C.-Y. Hsieh, Y.-H. Kao Yang, H.-J. Lin, C.-H. Chen, Y.-W. Chen, and Y.-H. Hu, "Developing a stroke severity index based on administrative data was feasible using data mining techniques," *Journal of Clinical Epidemiology*, vol. 68, no. 11, pp. 1292–1300, Nov. 2015.

[6] M. C. Paul, S. Sarkar, M. M. Rahman, S. M. Reza, and M. S. Kaiser, "Low cost and portable patient monitoring system for e-health services in bangladesh," in *2016 International Conference on Computer Communication and Informatics (ICCCI)*, 2016, pp. 1–4.

[7] S. M. Reza, M. M. Rahman, M. H. Parvez, M. S. Kaiser, and S. Al Mamun, "Innovative approach in web application effort & cost estimation using functional measurement type," in *2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*. IEEE, 2015, pp. 1–7.

[8] M. Asif-Ur-Rahman, F. Afsana, M. Mahmud, M. S. Kaiser, M. R. Ahmed, O. Kaiwartya, and A. James-Taylor, "Toward a heterogeneous mist, fog, and cloud-based framework for the internet of healthcare things," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4049–4062, 2018.

[9] H. M. Ali, M. S. Kaiser, and M. Mahmud, "Application of convolutional neural network in segmenting brain regions from mri data," in *International Conference on Brain Informatics*. Springer, 2019, pp. 136–146.

[10] M. Mahmud, M. S. Kaiser, A. Hussain, and S. Vassanelli, "Applications of deep learning and reinforcement learning to biological data," *IEEE trans. neural netw. learn. syst.*, vol. 29, no. 6, pp. 2063–2079, 2018.

[11] 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.

[12] S. M. Reza, M. M. Rahman, and S. Al Mamun, "A new approach for road networks-a vehicle xml device collaboration with big data," in *2014 International Conference on Electrical Engineering and Information & Communication Technology*. IEEE, 2014, pp. 1–5.

[13] C.-A. Cheng, Y.-C. Lin, and H.-W. Chiu, "Prediction of the prognosis of ischemic stroke patients after intravenous thrombolysis using artificial neural networks," *Studies in Health Technology and Informatics*, vol. 202, pp. 115–118, 2014.

[14] S. Cheon, J. Kim, and J. Lim, "The Use of Deep Learning to Predict Stroke Patient Mortality," *International Journal of Environmental Research and Public Health*, vol. 16, no. 11, 2019.

[15] M. S. Zulfiker, N. Kabir, A. A. Biswas, P. Chakraborty, and M. M. Rahman, "Predicting students' performance of the private universities of bangladesh using machine learning approaches," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 3, 2020.

[16] S. Rahman, T. Sharma, S. Reza, M. Rahman, M. Kaiser *et al.*, "Pso-nf based vertical handoff decision for ubiquitous heterogeneous wireless network (uhwn)," in *2016 International Workshop on Computational Intelligence (IWCI)*. IEEE, 2016, pp. 153–158.

[17] M. S. Singh and P. Choudhary, "Stroke prediction using artificial intelligence," in *2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON)*, Aug. 2017, pp. 158–161.

[18] C. Chin, B. Lin, G. Wu, T. Weng, C. Yang, R. Su, and Y. Pan, "An automated early ischemic stroke detection system using CNN deep learning algorithm," in *2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST)*, Nov. 2017, iSSN: 2325-5994.

[19] 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, pp. 1953–1959, Nov. 2018.

[20] T. Kansadub, S. Thammaboosadee, S. Kiattisin, and C. Jalayondeja, "Stroke risk prediction model based on demographic data," in *2015 8th Biomedical Engineering International Conference (BMEiCON)*, Nov. 2015, pp. 1–3.

[21] S. Y. Adam, A. Yousif, and M. B. Bashir, "Classification of Ischemic Stroke using Machine Learning Algorithms," *International Journal of Computer Applications*, vol. 149, no. 10, pp. 26–31, Sep. 2016.

[22] H. Lee, E.-J. Lee, S. Ham, H.-B. Lee, J. S. Lee, S. U. Kwon, J. S. Kim, N. Kim, and D.-W. Kang, "Machine learning approach to identify stroke within 4.5 hours," *Stroke*, vol. 51, no. 3, pp. 860–866, 2020.

[23] T. Kansadub, S. Thammaboosadee, S. Kiattisin, and C. Jalayondeja, "Stroke risk prediction model based on demographic data," in *2015 8th Biomedical Engineering International Conference (BMEiCON)*. IEEE, 2015, pp. 1–3.