ISSN: 2321-2152

IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN2321-2152

www.ijmece .com

Vol 12, Issue 1, 2024

DETECTION OF STROKE DISEASE USING MACHINE LEARNING ALGORITHMS

Ms. G.ANITHA, Assistant Professor, Department Of ECE SICET, Hyderabad
Rohini Reddy Goureddy, Naveen Duggenaboina, Vijay Chinnakotla, Laxmikala Dhanavath
UG Student, Department Of ECE, SICET, Hyderabad

ABSTRACT

A stroke is a medical condition in which cell death occurs due to insufficient blood flow to the brain. It is currently the leading cause of death worldwide. Several risk factors thought to be related to the cause of stroke have been discovered by testing affected individuals. Many studies have been conducted to predict and classify stroke diseases using these risk factors. Most of the models are based on data mining and machine learning algorithms. This study used four machine learning algorithms to detect the types of strokes that a person may have or have had, based on the person's physical condition and medical report data. We collect numerous inputs from hospitals and use them to solve problems. The classification results show that the results are satisfactory and can be used for real-time medical reports. We believe that machine learning algorithms can help improve the understanding of diseases and can be a good partner for healthcare. Index Terms - Stroke, Machine Learning, WEKA, Simple Bayes, J48, k-NN, Random Forest.

INTRODUCTION

A stroke occurs when cells die due to insufficient blood flow to the brain. The two main types of stroke are ischemic stroke and hemorrhagic stroke. Ischemic stroke occurs due to insufficient blood flow and hemorrhagic stroke occurs due to bleeding [1]. Another type of stroke is a transient ischemic attack. There are two types of ischemic stroke: embolic stroke and thrombotic stroke. An embolic stroke occurs when a blood clot forms in any part of the body and travels to the brain, blocking blood flow. Thrombotic stroke caused by a blood clot that reduces blood flow in an artery. Stroke is divided into two types: subarachnoid hemorrhage and intracerebral hemorrhage. Transient ischemic attacks are also known as “mini-strokes” [2].
Many people die from stroke and stroke is increasing in developing countries [3]. There are several risk factors for stroke that control different types of stroke. Predictive algorithms can help understand the relationship between these risk factors and stroke types. Machine learning algorithms can improve patient health through early diagnosis and treatment. W

Singh and Chaudhary developed a model using artificial neural networks (ANN) to predict stroke [6]. They collected their dataset from the Cardiovascular Health Study (CHS) database. Three datasets including 212 stroke cases (all three) and 52, 69 and 79 non-stroke cases were made. The final dataset contains 357 features and 1824 entities with 212 hits. During feature selection, the C4.5 decision tree algorithm was used, and principal component analysis (PCA) was used for dimensionality reduction. ANN implementation uses Back Propagation learning method. The accuracies of 95%, 95.2% and 97.7% were obtained for three data sets, respectively.

Adam et al developed a classification model for ischemic stroke using decision tree and nearest neighbor (k-NN) algorithm [7]. Their dataset was collected from several hospitals and medical centers in Sudan and is the first

ISSN2321-2152

www.ijmece .com

Vol 12, Issue 1, 2024

dataset on ischemic diseases in Sudan. It contains information of 15 characteristics and 400 patients. The experimental results show that the decision tree classification performance is higher than the k-NN algorithm.

Soda et al used decision trees, Bayesian classifiers and neural networks to classify stroke [8]. Their dataset contains 1000 records. PCA algorithm was used to reduce the dimensions. More than 10 rounds of each algorithm, neural network, naive Bayes classifier, and decision tree algorithm achieved the highest accuracy of 92, 91, and 94%, respectively.

Some methods such as [4] and [7] use very small data sets. Govindarjan et al. [2] predicted only two categories of stroke. Therefore, we proposed a method that uses a large dataset with four hit classes.
III. The formula of the problem
a. data source
We constructed the dataset by collecting stroke data from various sources. Our dataset includes patient information for a total of 1058 patients, of which 412 are male and 646 are female. Stroke types were reported as ischemic stroke in 437 cases, hemorrhagic stroke in 302 cases, small stroke in 142 cases, and stroke class in 177 cases. Although the data set is not perfectly symmetrically distributed,

METHODOLOGY

TABLE I: List of attributes of the dataset.

Sl.	Attributes	Description
1	Age	Age of the patient
2	Sex	Sex of the patient
3	Confusion	Health confusion
4	Vision Loss	Decreasing ability to see
5	Dizziness	A range of sensations, such as feeling faint, woozy, weak or unsteady
6	Headache	Symptom of pain anywhere in the region of the head or neck
7	Weaknessnausea	Feeling queasy or queasy in the stomach
8	Nausea	A sensation of unease and urge to vomit
9	Vomiting	Vomiting is the involuntary emptying of stomach contents through the mouth
10	Seizures	A seizure is a sudden, uncontrolled anxiety in the brain.
11	Loss of Balance	Loss the balancing sensation
12	Irregular Heartbeat	A situation when the heart beats too fast, slow, or irregularly.
13	Chest Discomfort	Feeling pressure or squeezing in the chest.
14	Fainting	Fainting is loss of consciousness caused by decreased blood flow to the brain
15	Fatigue	Fatigue is a feeling of constant tiredness or weakness
16	Difficulty Breathing	Feeling difficulty in breathing
17	Difficulty Speaking	Feeling difficulty in speaking
18	Hearing Loss	Reducing ability to hear.
19	Paralysis	Paralysis is the loss of function of muscle in any part of the body

ISSN2321-2152

www.ijmece .com

Vol 12, Issue 1, 2024

20	Sensation Loss	Being unable to feel pain, heat, or cold
21	CT	Computed Tomography result of the patient
22	CTA	Computed Tomography Angiography result of the patient
23	MRI	Magnetic Resonance Imaging result of the patient
24	CTP	Computer-To-Plate result of the patient
25	MRA	Magnetic Resonance Angiogram result of the patient
26	X-RAY	X-RAY result of the patient
27	ECG	Electrocardiogram result of the patient
28	ECO	Echocardiogram result of the patient

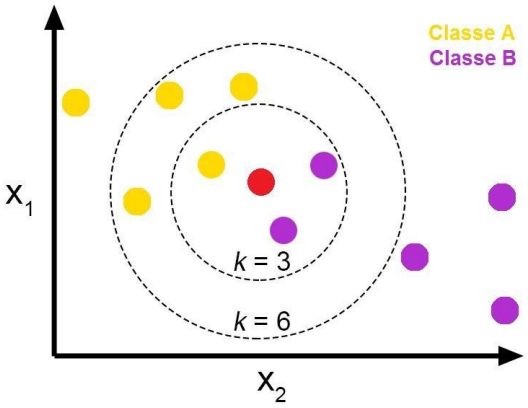


Fig. 1: k-NN.

0 is "woman", 1 is "female", etc. Some elements are missing from the dataset. Some attributes do not a pply to humans, that is, they are invalid. We replace them with zeros - 0 - to avoid exceptions. We also r emoved unnecessary words like "3 times" and made vomiting only 3 times etc. We replaced it with . Exa mples of preliminary data are shown in Table 2. Data Analysis
The Waikato Environment for Information Analysis (WEKA) is a machine learning tool developed and m anaged by the University of Waikato in New Zealand [10]. Previous studies

Show that WEKA is an excellent learning machine. Many similar projects have been completed with this method

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 1, 2024

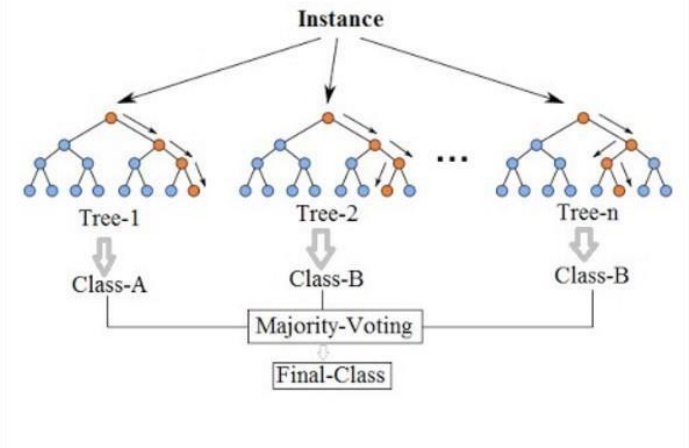
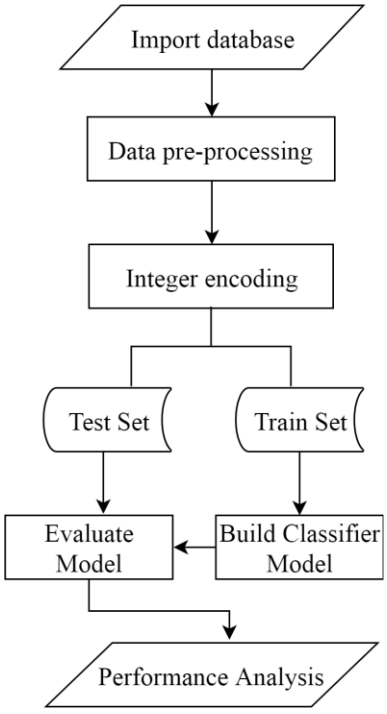


Fig. 2: Random Forest

weka and found it good [11] [12] [13] [14]. We use methods such as Naive Bayes, Random Forest and J 48 developed at WEKA for stroke diagnosis. These algorithms have been described before. First, we im port the data from the contour file. We use WEKA to classify hits after pre-
processing and integer encoding. Hit detection in WEKA performs the following steps:
Preliminary data and visualization
Character selection
Evaluation and training set segmentation
Classification using different algorithms



ISSN2321-2152

www.ijmece .com

Vol 12, Issue 1, 2024

Fig. 3: Work-flow of data mining .

- Model evaluation

The work-flow of data mining is given in Fig. 3.

EXPERIMENTAL RESULTS

To evaluate the performance, we have used Accuracy, Precision, Recall, and F1-score. Classification accuracy is the ratio of correct predictions to total number of predictions made by the model. Precision is the ratio of true positive to the true positive and false positive prediction. Recall is defined as the ratio of true positives to the true positive and false negative. F1- score or F-measure is the balance measure to express the performance in a single quantity. It is the harmonic mean of precision and recall They are formulated as follows:

(2)

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

(3)

$$Precision = \frac{TP}{TP + FP}$$

(4)

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

(5)

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Where, TP: correct positive prediction, FP: incorrect positive prediction, TN: correct negative prediction, FN: incorrect negative prediction, P: TP+FP, N: TN+FN. The confusion matrix for calculating TP, FP, TN, FN is given in Fig. 4.

TABLE II: Data pre-processing.

Attributes	Before processing	After processing
Age	30	30
Sex	Male	0
Confusion	POSITIVE	1
Vision Loss	central vision loss	4
Dizziness	POSITIVE	1
Headache	POSITIVE	1
Weaknessnausea	POSITIVE	1
Nausea	NEGATIVE	0
Vomiting	3 times	3
Seizures	NEGATIVE	0
Loss of Balance	POSITIVE	1
Irregular Heartbeat	NEGATIVE	0
Chest Discomfort	NEGATIVE	0

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 1, 2024

Fainting	NEGATIVE	0
Fatigue	POSITIVE	1
Difficulty Breathing	NEGATIVE	0
Difficulty Speaking	N/A	0
Hearing Loss	N/A	0
Paralysis	N/A	0
Sensation Loss	N/A	0
CT	POSITIVE	1
CTA	POSITIVE	1
MRI	POSITIVE	1
CTP	N/A	0
MRA	N/A	0
X-RAY	deformaties in the skull	2
ECG	N/A	0
ECO	N/A	0

Class	Accuracy	Precision	Recall	F-Measure
ISCHEMIC STROKE	1	1	1	1
HEMORRHAGIC STROKE	1	0.993	1	0.997
MINI STROKE	0.993	1	0.993	0.996
BRAIN STEAM STROKE	0.993	1	0.994	0.997

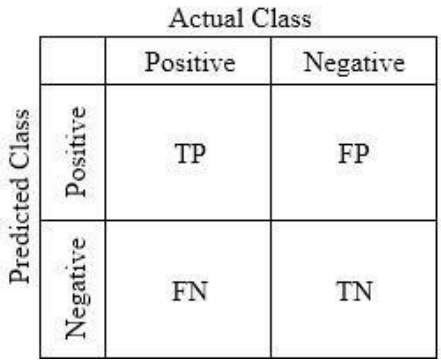


Fig. 4: Confusion matrix.

We have used a 10-fold cross validation for each algorithm. Performance comparison of different algorithms are shown in Table III.

From Table III, we see that the accuracy of Naive Bayes classifier is 85.6%. The accuracy for J48, k-NN and Random Forest is 99.8%. Naive Bayes has got the precision, recall, and f-measure as 88.1%, 85.6%, 86.1%. All of the J48, k-NN and TABLE III: Performance comparison of different algorithms.

ISSN2321-2152

www.ijmece .com

Vol 12, Issue 1, 2024

AlgorithmAccuracy Precision Recall F-Measure

Naive Bayes	0.856	0.881	0.856	0.861
J48	0.998	0.998	0.998	0.998
k-NN	0.998	0.998	0.998	0.998
Random Forest	0.998	0.998	0.998	0.998

- a = ISCHEMIC STROKE
- b = HEMORRHAGIC STROKE
- c = MINI STROKE
- d = BRAIN STEAM STROKE

Random Forest has the precision, recall, and f-measure same as 99.8%, 99.8%, and 99.8% respectively.. Detailed results for each class on every algorithm are shown in Table IV, V, VI, and VII.

TABLE IV: Detailed performance of Naive Bayes algorithm.

Class	Accuracy	Precision	Recall	F-Measure
ISCHEMIC STROKE	0.872	0.995	0.872	0.929
HEMORRHAGIC STROKE	0.801	0.913	0.801	0.854
MINI STROKE	0.803	0.74	0.803	0.77
BRAIN STEAM STROKE	0.955	0.66	0.955	0.781

Table IV shows that the Brain Stem stroke class gets a better classification result for the Naive Bayes classifier in terms of accuracy. In terms of F-measure, it is Ischemic stroke class.

TABLE V: Detailed performance of J48 algorithm.

TABLE VI: Detailed performance of k-NN algorithm.

Class	Accuracy	Precision	Recall	F-Measure
ISCHEMIC STROKE	0.998	1	0.998	0.999
HEMORRHAGIC STROKE	0.997	1	0.997	0.998
MINI STROKE	1	0.993	1	0.996
BRAIN STEAM STROKE	1	0.994	1	0.997

ISSN2321-2152

www.ijmece .com

Vol 12, Issue 1, 2024

Table VI and VII also report that k-NN (with Euclidean distance) and Random Forest classifiers have the highest level of classification results achieved so far in our models.

Confusion matrix for each individual algorithms is shown in Table VIII, IX, X, and XI. The classes are:

TABLE VII: Detailed performance of Random Forest algorithm.

Class	Accuracy	Precision	Recall	F-Measure
ISCHEMIC STROKE	0.998	1	0.998	0.999
HEMORRHAGIC STROKE	0.997	1	0.997	0.999
MINI STROKE	1	0.993	1	0.996
BRAIN STEAM STROKE	1	0.994	1	0.997

Naive Bayes	0.856	0.881	0.856	0.861

- a = ISCHEMIC STROKE

ISSN2321-2152

www.ijmece .com

Vol 12, Issue 1, 2024

TABLE VIII: Confusion matrix for Naive Bayes algorithm.

	a	b	c	d
a	381	10	29	17
b	2	242	4	54
c	0	12	114	16
d	0	1	7	169

TABLE IX: Confusion matrix for J48 algorithm.

	a	b	c	d
a	437	0	0	0
b	0	302		
c	0	1	141	
d	0	1	0	176

stroke can be done by adding some non-stroke data with the existing dataset.

TABLE X: Confusion matrix for k-NN algorithm.

	a	b	c	d
a	436	0	0	1
b	0	301	1	0
c	0	0	142	0
d	0	0	0	177

Naive Bayes is a very simple classifier, so you shouldn't expect it to be more powerful. From the analysis of classification results, it can be said that J48, k-NN and random forest did their job well in diagnosing stroke diseases.

CONCLUSION

In this paper, a sufficiently large dataset of stroke patients was accurately classified. Four classifiers as shown in Table 11: Confusion matrix for random forest algorithm.

	a	b	c	d
a	436	0	0	1
b	0	301	1	0
c	0	0	142	0
d	0	0	0	177

Naive Bayes, J48, k-NN and random forest were used to diagnose stroke. Performance analysis shows that Naive Bayes performs better than other techniques. The novelty and main contribution of our work is collecting this dataset and preparing it for use in WEKA. This model helps to show warnings that people are having a stroke. The healthcare industry generates vast amounts of complex data about patients, hospital resources, diagnoses, electronic patient records, medical devices, and more. Correlation of these data is very difficult even for experts in the field. This helps doctors better understand the type of disease. A limitation of our method is that the dataset is not perfectly symmetric. However, it did not affect the prediction accuracy of other algorithms. The Naive Bayes algorithm did not work as expected.

Future research can extend the study by using different classification techniques. In addition, the forecast is below

REFERENCES

[1] S. H. Pfaffs, A. T. Hansen, and A.-M., "Evaluation of thrombophilia in young patients with stroke," Thrombosis, vol. 137,
Pages 108–112, 2016.
[2] P. Govindarajan, R. K. Soundarapandian, A. H. Gandomi, R. Patan, P. Jayaraman, and R. Manikandan, "Stroke Disease Classification Using Machine Learning Algorithms," Neurocomputing and Applications, pp. 1- 12.
[3] L. T. Kohn, J. Corrigan, M. S. Donaldson, et al., torr is human: Building a security health system, vol 6. National Academy Press, Washington, DC, 2000.
[4] R. Jeena and S. Kumar, "Stroke prediction using svm", 2016 International Conference on Control, Measurement, Communication and Computing Technologies.

[5] P. A. Sandercock, M. Niewada, and A. Członkowska, "International Stroke Clinical Database," Trials, vol 13, no. 1, pp. 1-1, 2012.
[6] M. S. Singh and P. Choudhary, “Stroke prediction using artificial intelligence”, 2017 8th Industrial Automation and Electromechanical Engineering Conference (IEMECON), pp. 158-161, IEEE, 2017.
[7] S. Y. Adam, A. Yousif, and M. B. Bashir, “Ischemic stroke classification using machine learning algorithms,” Int J Comput Appl, vol. 10, pp. 26-31, 2016.
[8] A. Sudha, P. Gayathri and N. Jaisankar, "Effective analysis and prediction models of stroke diseases using classification methods," International Journal of Computer Applications, vol 43, no. 14, pp. 26-31, 2012.
[9] G. Kaur and A. Chhabra, "An improved j48 classification algorithm for diabetes prediction," International Journal of Computer Applications, vol.98, no. 22, 2014.
[10] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufman, 2016.
[11] P. Sewaiwar and K. K. Verma, "A comparative study of decision tree classification algorithms using weka", International Journal of Emerging Research in Management and Technology, Volume 4, pp. 2278-9359, 2015.
[12] K. A. Shakil, S. Anis, and M. Alam, “Prediction of dengue fever using weka data mining tools”, arXiv preprint arXiv:1502.05167, 2015.
[13] J. A. Alkurimi, H. A. Jarab, L. E. George, A. R. Ahmad, A. Saliman,
K. Al-Jashmy, “A comparative study using weka for red blood cell classification,” International Journal of Medicine, Health, Pharmacy and Biomedical Engineering, vol. 9, no. 1, pp. 19-22, 2015.
[14] M. S. Siddiqui and A. I. Abidi, "Comparative study of different classification techniques using weka tool," Global Sci-Tech, vol. 4, pages 200-208, 2018.