JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

ROLE OF AI APPROACHES IN STROKE PREDICTION – A REVIEW

[1]Janvi Shah, [2]Mittal Darji, [3]Vinita Shah

[1]Research Scholar, [2]Assistant Professor, [3] Assistant Professor
Department of Information Technology,
G H Patel College of Engineering & Technology, Anand, India

Abstract : Stroke is a serious life-threatening disease that damages the brain, just like a heart failure that damages the heart. It is third leading cause of death. Prediction of stroke is time consuming and frustrating for doctors. Chances of survival are more when treatment begins quickly. We can observe symptoms like sudden numbness or weakness within the face, arm or leg, especially on one side of the body, sudden confusion, trouble speaking, or difficulty understanding speech, sudden trouble seeing in one or both eyes. Various machine learning algorithms like SVM, RFA, DTA, Logistic Regression and KNN is employed. In this paper we have discussed about various machine learning algorithms which can be used to predict stroke. Comparative analysis of algorithms is also discussed.

IndexTerms **- Stroke prediction, Machine learning approaches, Artificial Intelligence, Random Forest.**

I. INTRODUCTION

Stroke could be a serious life-threatening disease that damages the brain, just like a heart failure that damages the heart. It doesn't circulate blood and there's not enough oxygen within the brain cells. Stroke can cause loss of mobility, sudden hurting, speech impotence, amnesia and limited thinking, confusion or death. Stroke affects adults of all ages. It will be managed through beneficial regulation and it's important to correct the chance factors. Stroke prediction could be a workforce and a run of the mill for doctors. Hypertension and Hyperlipidemia are a standard stroke risk factor.

The machine learning model receives patient information and suggests appropriate expectations. The platform can remove hidden information from enrolled clinical datasets, predict infected patients and use clinical profile such as age, blood pressure, blood sugar levels, etc. to predict the likelihood of a patient becoming ill. The clinical record additionally includes the patient's medical and stroke history, assuming they have had a previous stroke and taking all these data to train the machine according to different models.

Types of Stroke:

In medical terms, there are basically three sorts of stroke. The primary is an ischemic stroke, a type of stroke caused by blood clotting in the brain. The second is hemorrhagic stroke due to bleeding. The third could be a transient ischemic stroke, also referred to as a mini-stroke, which occurs when the blood supply to the brain is discontinued. In this paper, we use predictive classification algorithms like decision trees, SVM, logistic regression, and naive compartments for stroke prediction and principle component analysis algorithms for attribute reduction. During this paper, we use predictive classification algorithms such as decision trees, SVM, logistic regression, and naive compartments for stroke prediction and principle component analysis algorithms for attribute reduction.

II. LITERATURE SURVEY

A comparative study of five different models showed that Random Forest, Logistic Regression, and K nearest neighbors are 95.5% accurate, while Decision Trees are 91.13% accurate and Support Vector Machine 92.43% accurate. Finally, Random Forest was chosen because the best model with high accuracy and low false negatives. [1]

In 2018, authors discussed the employment of AI in stroke prediction. It also uses inspiring algorithms and structures that include patient characteristics like gender, age, height, BMI, etc. and build data models with decision tree algorithms to research these parameters. The results were analyzed using the error classification table and therefore the accuracy was 95%. To the current end, we created a training model that helps us compare our research data. And supported this comparison, a report was created. [2]

In 2012, a stroke model using the classification method(PCA) was proposed by authors. It uses Decision Tree, NaiveBayes and Neural Network Editing algorithms to predict impact types with reasonable quality. Reduce the size using part of a software analysis algorithm (PCA). The Bayesian classifier has reached 87.10% and 91.30% respectively. Theycompared these strategies and choose solutions to best category. The proposed model is captured by capturing patient information and using properties reduction. Accuracy is measured based on sensitivity and clarity. The functionality of the neural network was that it was more accurate than the other two classification methods. [3]

In 2010, authors proposed a new algorithm for automatic selection of the ability to sort popular properties based on proposed rules. They combined with a vector vehicle aimed at major areas under ROC curve (AUC). They also proposed a boundary inversion algorithm with a line separator to find a more similar indicator than a COX model. This model is used for medical predictions and other diseases that are distributed and therefore the results don't seem to be fully understood. However, they concluded that this feature of the algorithm's choice canwork well within the other datasets with highly mentioned functions, like evaluating the performance of each element individually. To beat this problem, use a standard L1 feature complementation algorithm to minimizefeatures before applying conservative fine-tuning componentselection. [4]

This new approach to automated system development topredict the stroke in 2019 is a new approach to the development of the results, including the consistency, high accuracy and fast reporting forecasts. Additionally, since theseprediction algorithms might have several task steps, you can adjust sensitivity and specificity according to clinical DNN requirements, so you can use small datasets to get the best performance than the GBDT method. Additional Research Determines the practicality to apply DNN in a clinical environment and it is needed to check whether the useof DNN can be determined whether or not the results of clinical nursing and patients can be improved. [5]

III. Machine Learning Approaches In Stroke Prediction

There are many machine learning algorithms like regression toward the mean, logistic regression, decision tree, SVM, naïve bayes algorithm, KNN algorithm, K-Means algorithm, and random forest algorithm. These machine learning algorithms play a really major duty in healthcare. Among these algorithms, five methods are mainly used for stroke prediction: Random Forest, Decision Tree, Logistic Regression and K-Nearest Neighbor (KNN).

First, Random Forest might be a popular machine learning algorithm connected with supervised learning techniques. It is used for classification and regression problems in machine learning.A classifier that includes multiple decision trees for different subsets of a given dataset and extracts a mean to boost the prediction accuracy of that dataset. The more trees within theforest, the higher the accuracy and avoidance of the overfitting problems. The diagram below illustrates how thisalgorithm works:

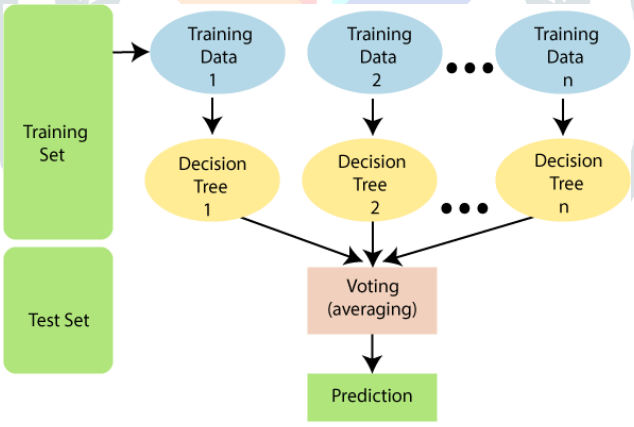

Figure 1 – Random Forest Algorithm

Random forest algorithm requires less training time and predicts output with high accuracy compared to other algorithms. It works efficiently even on large datasets. It canalso maintain accuracy in most cases the information ismissing. The workflow can be illustrated within thesubsequent steps:

Step 1: Choose the random k data points from the training set.
Step 2: Construct a call decision tree related to the selected data points.
Step 3: Select the quantity N for the choice tree you would like to create.
Step 4: Repetition of steps 1 and a couple of.
Step 5: Considering the new data, search the predictions in every decision tree and allocate the new data point to the very best voting group.

Second technique which is Decision Tree which can be used for classification and Regression problems, but majorly preferred for solving Classification problems. It is a tree-structured classifier in which internal nodes constitutes the features of a dataset, branches constitutes the selection rules andeach leaf node represents the top result. In a decision tree there are two nodes which are the selection node and leaf node.Decisions nodes are want to form any decision and have multiple branches, whereas Leaf nodes are the output of thosedecisions and don't contain to any extent further branches. Achoice tree simply asks a matter and supported the answer (Yes/No); it further grouped into sub trees. Below diagram explain the overall structure of a call tree:

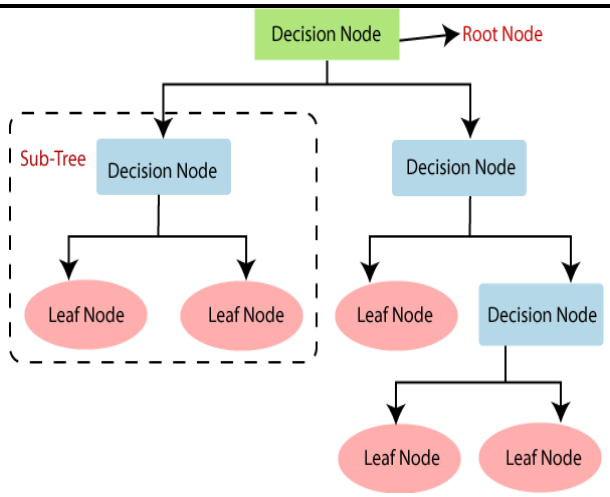

Figure 2 – Decision Tree Algorithm

There are the two purposes for using the decision trees: decision tree is mostly easy to know by copying humanthinking ability when making decisions and the reasoning of thedecision tree is easy to recognize because it shows the tree structure.

The following algorithm gives you a more robust recognizing of whole process.
Step-1: Start the tree at the foundation node containing thewhole dataset.
Step-2: Use Attribute Selection Measure (ASM) to search outthe simplest attribute in the dataset.
Step-3: Split the S into subsets including possible valuesfor the simplest attributes.
Step-4: Construct a decision tree node containing the best attribute.
Step-5: Iteratively construct a new decision tree using thesubsets of the dataset created in step -3.
Continue this process until you reach a stage where you will further classify the nodes and name the last node as final node.

The next approach is Logistic Regression machine learning algorithms; want to predict a categorical quantity employing a given set of independent variables. Predict the output of a categorical variable. It's going to be either Yes or No, 0 or 1, true or false, etc. but instead of giving the precise value as 0 and 1, it gives probabilistic values that lie between 0 and 1. Logistic regression is utilized for solvingthe classification problems.

Logistic Regression may be a foremost machine learning algorithm because it provides probabilities and allows you to classify new data. Logistic Regression is also acquainted with classify the observations using different kinds of data and it makes it easy to figure out which variables are the foremost effective for the classification. The image below shows the logistic capabilities:

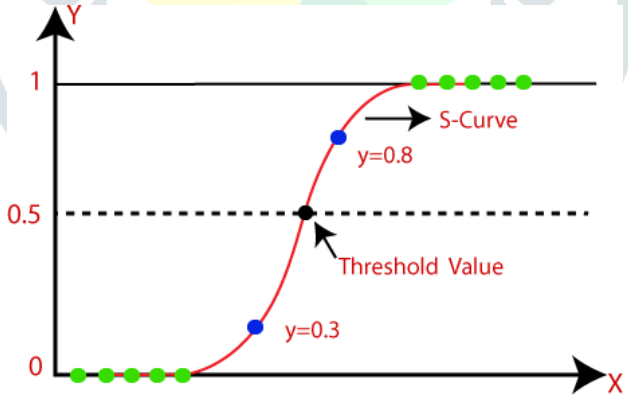

Figure 3 – Logistic Regression

Sigmoid Function: A sigmoid function is also an l function familiar with compare predicted values to probabilities. Thevalues of the logistic regression should between 0 and 1, andcan't exceed this limit, thus forming a curve just like the "S" form. S-form curve a bit like the sigmoid function or the logistic function. Logistic regression uses the concept of the edge to define the probability of either 0 or 1.

The latter way, K-Nearest Neighbor is one in every of the sole algorithms for training machines supported controlled learning technologies. KNN algorithm includes new cases/data and affordable similarities and puts new examples within the favorite categories where the available categories are available. KNN algorithm stores all the available data andclassifies a replacement data points supported similarity. In other words, it implies that new data is displayed and you willbe able to easily categorize with a luxury category in an exceedingly well using the KNN algorithm. The KNN algorithm isn't only classified yet as multivariate analysis, butis particularly used for classification operations. It's also a lazystudent algorithm. Because it maintains the research instantly within the training session, it immediately maintains the knowledge and performs action on the data set during classification.

The last method is Support Vector Machine (SVM) is one amongst the foremost popular supervised learning algorithms, which is employed for both Classification and Regression problems. However, it's mainly used for classification problems in machine learning. The goal of the SVM algorithm is to make the simplest line or decision line or boundary which will divide n-dimensional space into classes making it easy to put the new data point into the proper categories within the future. SVM picks the tip points/vectors that help to form the hyper plane. Since these edge cases are called as support vectors, and also the algorithm is termed as Support Vector Machine. Consider the diagram below with two categories classified employing a decision boundaries or hyper planes.

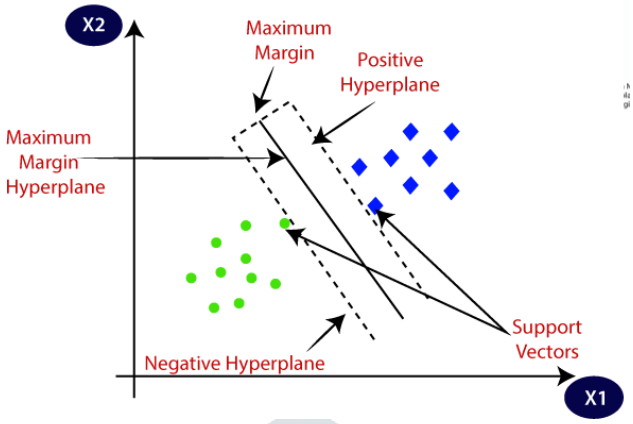

Figure 4 - SVM

IV. Comparative Analysis

Harshitha K V et al. [1] proposed an acute stroke model using Decision tree, logistic regression, random forest, support vector machine and KNN algorithms were used but random forest, logistic regression and K-nearest neighbor gives 95.5% accuracy, whereas decision tree accuracy was 91.13%, The support vector machine gives 92.43% accuracy.

Aishwarya Roy et al. [2] used Decision Trees, Bayesian Classifier and Neural Networks which train the model. The accuracy of decision tree algorithms and KNN were 95.42% and 94.18%, in stroke prediction. Results were employing a confusion matrix.

Sudha et al. [3] observed models using decision trees, KNN, SVM, Neural Network, logistic regression, random forest, naïve Bayes. The random forest model showed the simplest results with average values of 0.93+0.03 and 0.9±0.01 respectively. They compared these methods and chose the decision tree as the best classification Method.

Aditya Khosla et al. [4] proposed a model that may help doctors in clinical trials. Logistic Regression with random forest showed the very best accuracy of deep neurological networks.

Chen-Ying Hung et al. [5] has provided models supported gradation enhancement of deep neural network (DNN), gradient boosting decision tree (GBDT). In this study, AUC of 92% promotions are achieved with DNN and GBDT algorithm where, DNN requires fewer educational data. This paper demonstrates that DNN could also be a promising model method.

V. Conclusion

We can use the foremost risk factors to predict stroke using deep learning algorithm. For better accuracy, systolic blood pressure, diastolic pressure level, low blood pressure, moderate blood pressure, low, high values and heart rate attributes can be appraised. However, the random forest is considered one of the main powerful and approved survey strategies. Hybrid approach of any two machine learning algorithms or deep learning algorithm can be employed for better accuracy.

References

[1] [1] Harshitha K V, Harshitha P, Gunjan Gupta Vaishak P, Prajna K B, "Stroke Prediction Using Machine Learning Algorithm", International Journal of Innovative Research in Engineering & Management (IJIREM), Volume-8, Issue-4, July 2021

[2] Hamed Asadi, Richard Dowling and Bernard Yan, "Machine Learning for outcome prediction of acute ischemic stroke", PLOSONE Vol.9 Issue2, Feb2014.

[3] Aishwarya Roy, Anwesh Kumar, Navin Kumar Singh and Shashank D, "Stroke Prediction using Decision Trees in Artificial Intelligence", IJARIIT, Vol.4, Issue2, pp:1636-1642,2018.

[4] Chiun-Li-Chin, Guei-Ru Wu, Bing-Jhang Lin, Tzu-ChiehWeng, Cheng-Shiun Yang, Rui-CihSu and Yu-Jen Pan, "An Automated Early Ischemic Stroke Detection System using CNN Deep Learning Algorithm", IEEE 8th International Conference on Awareness Science and Technology, 2017.

[5] A. Sudha, P. Gayathri, "Effective analysis & predictive model of stroke disease using classification methods", IJCA (0975-8887), Vol.43-No.14, April 2012.

[6] Aditya Khosla, Yu cao, Honglak Lee & Associates, "An integrated machine learning approach to stroke prediction", 25-28 July 2010, Washington, DC, USA.

[7] Balar Khalid and Naji Abdelwahab, "A model for predicting Ischemic stroke using Data Mining algorithms", IJISET, Vol. 2 Issue 11, Nov 2015, ISSN: 2348-7968.

[8] Mrs. Veena Potdar, Mrs. Lavnya Santosh, Yashu Raj Godwa, "A Survey on Stroke Disease Classification and Prediction using Machine Learning Algorithms", IJERT 2021 International Journal of Engineering Research & Technology ISSN: 2278-0181 Vol. 10 Issue 08, August-2021.
[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436-44, May 28 2015.
[10] JoonNyung Heo , Jihoon G. Yoon , Hyungjong Park , Young Dae Kim , Hyo Suk Nam and Ji Hoe Heo. "Stroke prediction in acute stroke", Stroke. 2019;50:1263-1265, AHA Journal, 20 Mar 2019.
[11] Benjamin Letham,Cynthia Rudin Tyler, H. McCormick and David Madigan "An interpretable model for stroke prediction using bayesian analysis", The Annals of Applied Statistics 2015, Vol. 9, No. 3, 1350–1371, Institute of Mathematical Statistics, IEEE, 5 November 2015.
[12] H. Du, M. M. Ghassemi, and M. Feng, "The effects of deep network topology on mortality prediction," 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2602-2605, 2016.
Claims database. "In 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 3110-3113.IEEE, 2017.
[13] Ankitha S, Jahnavi V, Harshavardhan N, Deepthi M an Artificial Intelligence Approach for Predicting Different Types of Stroke (IJERT) ISSN: 2278-0181 Vol. 8 Issue 12, December-2019.
[14] Gangavarapu Sailasya1 , Gorli L Aruna Kumari2 Analyzing the Performance of Stroke Prediction using ML Classification Algorithms International Journal of Advanced Computer Science and Applications, (IJACSA) Vol. 12, No. 6, 2021.
[15] Revanth.S1, Sanjay.S2, Sanjay.N3, Vijayaganth. V, International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE) Stroke Prediction Using Machine Learning Vol. 9, Issue 5, May 2020.
[16] Youngkeun Choi 1 , Jae Won Choi 2 Stroke Prediction Using Machine Learning based on Artificial Intelligence volume 9, No.5,September-october 2020.
[17]B.P. Deepak Kumar , Sagar Yellaram, Sumanth kothamasu , SurendharReddy Puchakayala, Heart Stroke Prediction using Machine Learning, International Journal of Creative Research Thoughts (IJCRT) Volume 9, Issue 6 June 2021.
[18] Suraj Kumar Gupta, Aditya Shrivastava, S. P. Upadhyay, Pawan Kumar Chaurasia A Machine Learning Approach for Heart Attack Prediction International Journal of Engineering and Advanced Technology (IJEAT),Volume-10 Issue-6, August 2021
[19] Meghana R, Trisiladevi C. Nagavi* STROKE DISEASE TYPES PREDICTION USING MACHINE LEARNING APPROACHES International Journal of Scientific Research (IJSR) Volume - 9 Issue - 10 October - 2020
[20] Ms. Gagana M 1, Dr. Padma M C2 Stroke Type Prediction using Machine Learning and Artificial Neural Networks, International Research Journal of Engineering and Technology (IRJET) Volume: 08 Issue: 06,June 2021.