ISSN 2277-2685
IJESR/DEC 2021/ Vol-11/Issue-4/1-13
Mr.E.Sunil/ International Journal of Engineering & Science Research

STROKE DISEASE IDENTIFICTION SYSTEM BY USING DIFFERENT TYPES OF MACHINE LEARNING ALGORITHMS

Mr.E.Sunil,
Assistant Professor, Dept. of CSE,
Malla Reddy Engineering College (Autonomous), Secunderabad, Telangana State

ABSTRACT:

Machine learning (ML) is a part of artificial intelligence (AI) that makes software applications to gain the exact accuracy to predict the end results not having to be directly involved to get the work done. This review aims to identify and analyze the Machine Learning approaches used for Stroke Prediction. We have considered the previously published works to review the Machine learning techniques used for Stroke Predictions. It's been found that the majority of the research work was done on mortality rate and functional outcome as the predicted outcomes. The most commonly used techniques were random forest, support vector machines, decision trees and neural networks. However, a few predictors and classifiers did primitive reporting standards for medical sector tools and none of which proved to be of any practical use.

Keywords:AI, Stroke predictions, random forest, ML.

1. INTRODUCTION:

Many people fall victim to stroke and the numbers are increasing more in the developing countries. Several risk factors play a role in determining various types of stroke. Predictive algorithms establish a relationship between the risk factors and the types of strokes. Machine learning algorithms helps in early diagnosis and prevention of these stroke cases. It is very difficult to predict the stroke symptoms and outbreaks taking note on the risk factors, since stroke is a complicated medical condition. This has enhanced the interests of people in technology sector to apply machine learning techniques to diagnose

*Corresponding Author

the stroke effectively by routinely collecting the datasets and delivering the accurate results for diagnosis. Furthermore, many papers have been published frequently which explains machine learning techniques to address the issue. The agenda of this survey paper is to identify the better machine learning techniques used to predict stroke, which will also help to understand and resolve the problem in more effective ways.

2. LITERATURE SURVEY

1. Thrombophilia testing in young patients with ischemic stroke: The possible significance of thrombophilia in ischemic stroke remains controversial. We aimed to study inherited and acquired thrombophilias as risk factors for ischemic stroke, transient ischemic attack (TIA) and amaurosisfugax in young patients.

We included patients aged 18 to 50 years with ischemic stroke, TIA or amaurosisfugax referred to thrombophilia investigation at Aarhus University Hospital, Denmark from 1 January 2004 to 31 December 2012 (N = 685). Clinical information was obtained from the Danish Stroke Registry and medical records. Thrombophilia investigation results were obtained from the laboratory information system. Absolute thrombophilia prevalences and associated odds ratios (OR) with 95% confidence intervals (95% CI) were reported for ischemic stroke (N = 377) and TIA or amaurosis fugax (N = 308). Thrombophilia prevalences for the general population were obtained from published data.

2. Classification of stroke disease using machine learning algorithms:

This paper presents a prototype to classify stroke that combines text mining tools and machine learning algorithms. Machine learning can be portrayed as a significant tracker in areas like surveillance, medicine, data management with the aid of suitably trained machine learning algorithms. Data mining techniques applied in this work give an

overall review about the tracking of information with respect to semantic as well as syntactic perspectives. The proposed idea is to mine patients' symptoms from the case sheets and train the system with the acquired data. In the data collection phase, the case sheets of 507 patients were collected from Sugam Multispecialty Hospital, Kumbakonam, Tamil Nadu, India. Next, the case sheets were mined using tagging and maximum entropy methodologies, and the proposed stemmer extracts the common and unique set of attributes to classify the strokes. Then, the processed data were fed into various machine learning algorithms such as artificial neural networks, support vector machine, boosting and bagging and random forests. Among these algorithms, artificial neural networks trained with a stochastic gradient descent algorithm outperformed the other algorithms with a higher classification accuracy of 95% and a smaller standard deviation of 14.69.

3. Stroke prediction using svm:

Early diagnosis of stroke is essential for timely prevention and treatment. Investigation shows that measures extracted from various risk parameters carry valuable information for the prediction of stroke. This research work investigates the various physiological parameters that are used as risk factors for the prediction of stroke. Data was collected from International Stroke Trial database and was successfully trained and tested using Support Vector Machine (SVM). In this work, we have implemented SVM with different kernel functions and found that linear kernel gave an accuracy of 90 %.

4. The international stroke trial database:

The International Stroke Trial (IST) is one of the largest randomized trials ever conducted on individual patients in acute stroke. The IST dataset includes data on 19 435 patients with acute stroke, with 99% complete follow-up. Over 26.4% patients were aged over 80 years at study entry. Background stroke care was limited and none of the patients

received thrombolytic therapy. This clinical trial was conducted between 1991 and 1996 and a pilot phase between 1991 to and 1993. This study is a large, prospective, randomized controlled trial, with 100% complete baseline data and over 99% complete follow-up data. For each randomized patient, data were extracted on the variables assessed at randomization; the early outcome point was 14-days after randomization or prior discharge, and at 6-months and provided as an analyzable database. The aim of the trial was to establish whether early administration of aspirin, heparin, both or neither influenced the clinical course of an acute is chaemic stroke.

5. Effective analysis and predictive model of stroke disease using classification methods:

In today world data mining plays a vital role for prediction of diseases in medical industry. Stroke is a life threatening disease that has been ranked third leading cause of death in states and in developing countries. The stroke is a leading cause of serious, long term disability in US. The time taken to recover from stroke disease depends on patients‟ severity. Number of work has been carried out for predicting various diseases by comparing the performance of predictive data mining. Here the classification algorithms like Decision Tree, Naive Bayes and Neural Network is used for predicting the presence of stroke disease with related number of attributes. In our work, principle component analysis algorithm is used for reducing the dimensions and it determines the attributes involving more towards the prediction of stroke disease and predicts whether the patient is suffering from stroke disease or not.

Existing System:

A stroke is a medical condition in which poor blood flow to the brain results in cell death. It is now a day a leading cause of death all over the world. Several risk factors believe to

be related to the cause of stroke has been found by inspecting the affected individuals. Using these risk factors, a number of works have been carried out for predicting and classifying stroke diseases. Most of the models are based on data mining and machine learning algorithms. In this work, we have used four machine learning algorithms to detect the type of stroke that can possibly occur or occurred form a person's physical state and medical report data. We have collected a good number of entries from the hospitals and use them to solve our problem. The classification result shows that the result is satisfactory and can be used in real time medical report. Naive Bayes Algorithm, J48 Algorithm, KNN Algorithm, And Random Forest Algorithms in existing system.

Proposed System:

Stroke is an injury that affects the brain tissue, mainly caused by changes in the blood supply to a particular region of the brain. As consequence, some specific functions related to that affected region can be reduced, decreasing the quality of life of the patient. In this work, we deal with the problem of stroke detection in Computed Tomography (CT) images using Convolutional Neural Networks (CNN) optimized by Particle Swarm optimization (PSO). We considered two different kinds of strokes, ischemic and hemorrhagic, as well as making available a public dataset to foster the research related to stroke detection in the human brain. The dataset comprises three different types of images for each case, i.e., the original CT image, one with the segmented cranium and an additional one with the radiological density's map. The results evidenced that CNN's are suitable to deal with stroke detection, obtaining promising results.

3. METHODOLOGY AND IMPLEMENTATION

To implement this project we have designed following modules

1) Upload Stroke Dataset:

Using this module we will upload dataset to application

2) Dataset Preprocessing & Features Selection:

Using this module we will clean dataset by replacing missing values with 0 and then apply label encoding algorithm to convert non-numeric values to numeric values and then select features from dataset and then split dataset into train and test where application used 80% data for training and 20% for testing

3) Train Naive Bayes Algorithm:

Above training data will be input to Naïve Bayes algorithm to train a model and this model will be applied on test data to calculate accuracy

4) Train J48 Algorithm:

Above training data will be input to J48 algorithm to train a model and this model will be applied on test data to calculate accuracy

5) Train KNN Algorithm:

Above training data will be input to KNN algorithm to train a model and this model will be applied on test data to calculate accuracy

6) Train Random Forest Algorithm:

Above training data will be input to Random Forest algorithm to train a model and this model will be applied on test data to calculate accuracy

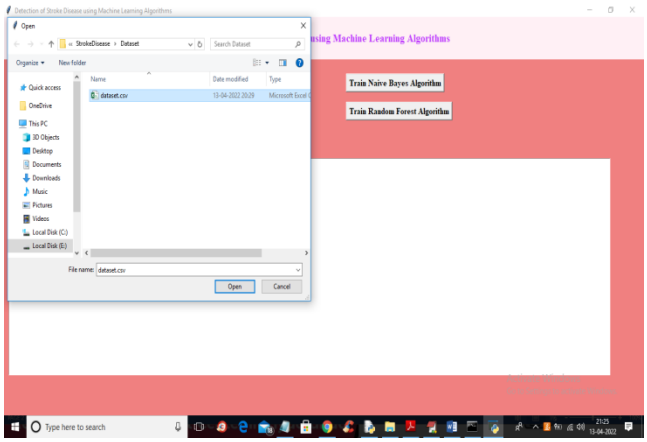
7) Train ANN Algorithm:

Above training data will be input to ANN algorithm to train a model and this model will be applied on test data to calculate accuracy

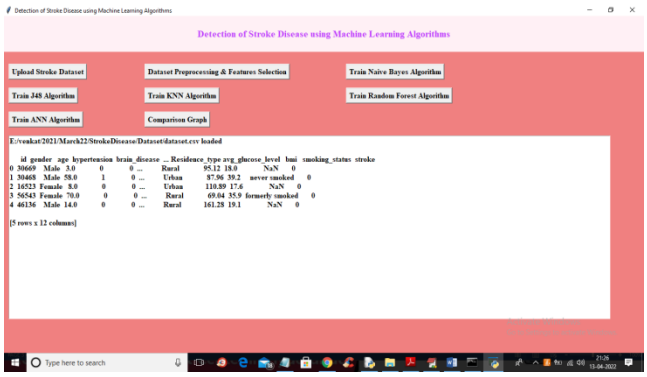
8) Comparison Graph:

Using this module we will plot accuracy comparison graph between all algorithms.

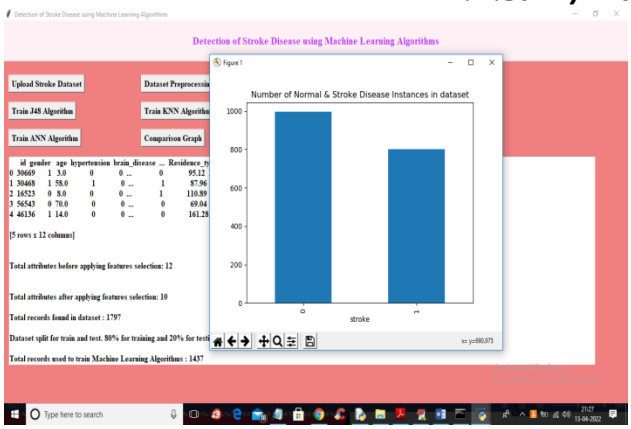
OPERATION:



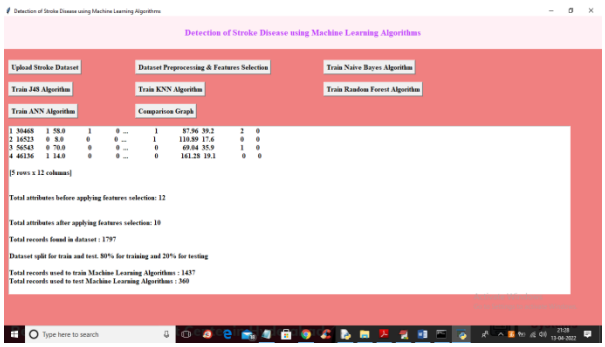
In above screen selecting and uploading dataset.csv file and then click on 'Open' button to load dataset and to get below output



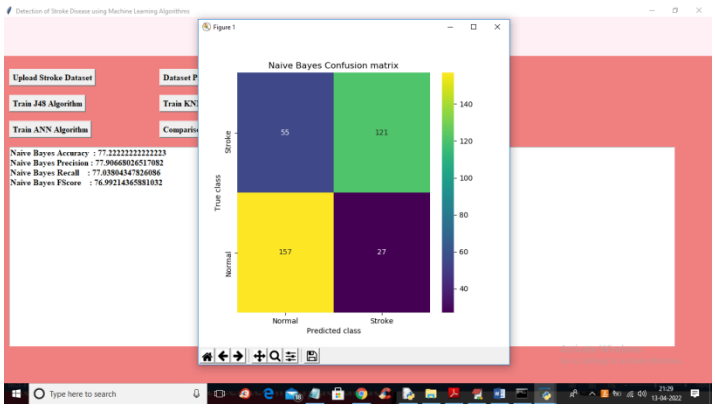
In above screen we can see dataset loaded and dataset contains so many missing and non-numeric data so click on 'Dataset Preprocessing & Features Selection' button to process dataset and to get below output



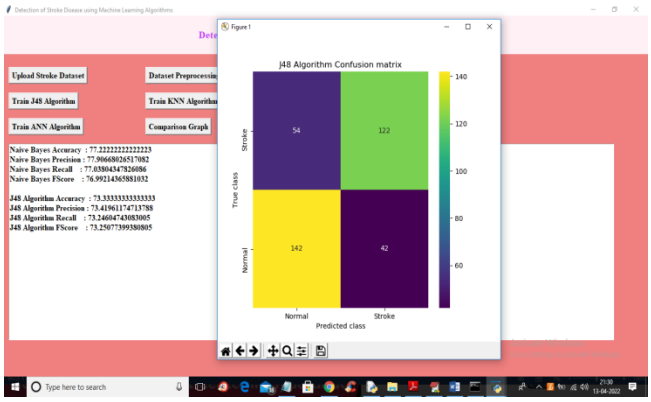
In above graph x-axis represents 0 (normal) and 1 (stroke) and y-axis represents number of instances available in those categories in dataset and now close above graph and see below screen



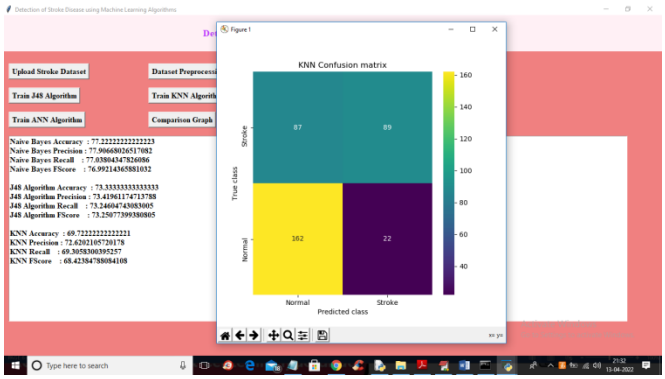
In above screen we can see all dataset converted to numeric format and then split dataset into train and test and now click on ‘Train Naïve Bayes Algorithm’ button to train Naïve Bayes on above dataset and get below output



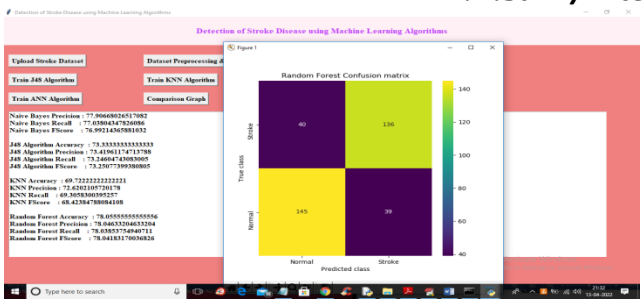
In above screen with Naïve Bayes we got 77% accuracy and in confusion matrix graph we can see number of correct and incorrect prediction by Naïve Bayes. Now click on 'Train J48 Algorithm' button to get below output



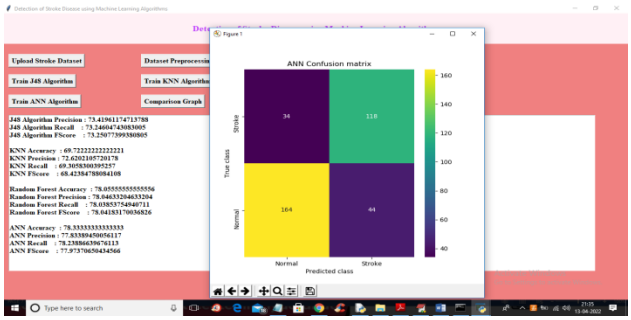
In above screen with J48 we got 73% accuracy and in confusion matrix graph we can see number of correct and incorrect prediction by J48.Now close above Graph and then click on 'Run KNN Algorithm' button to get below output



In above screen with KNN we got 69% accuracy and in confusion matrix graph we can see number of correct and incorrect prediction by KNN. Now close above Graph and then click on 'Run Random Forest Algorithm' button to get below output



In above screen with Random Forest we got 78% accuracy and in confusion matrix graph we can see number of correct and incorrect prediction by Random Forest. Now close above Graph and then click on 'Run ANN Algorithm' button to get below output



In above screen with ANN we got 78.33% accuracy and in confusion matrix graph we can see number of correct and incorrect prediction by ANN and in all algorithm ANN got high accuracy. Now close above Graph and then click on 'Comparison Graph' button to get below graph.

In above graph x-axis represents algorithm names and y-axis represents accuracy and other metrics like precision, recall etc. different colour bar represents different metrics and in all algorithms ANN got high accuracy.

CONCLUSION

Based on the result for classifying stroke from CT head scan image, ANN is able to help neurologist to classify stroke. The obtained accuracy also depends on the number of acquired data for training dataset. In this research, our proposed method can give 78.33% of accuracy for testing N images of each type of stroke. The classification result much depends on how much images that being used in training process. More images used in training process, the higher the accuracy. Here different colour bar represents different metrics and in all algorithms ANN got high accuracy.

In future work, it is possible to extend the research by using different classification techniques. Moreover, the prediction of stroke can be done by adding some non-stroke data with the existing dataset.

REFERANCES

[1] S. H. Pahus, A. T. Hansen, and A.-M. Hvas, “Thrombophilia testing in young patients with ischemic stroke,” Thrombosis research, vol. 137, pp. 108–112, 2016.

[2] 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, pp. 1–12.

[3] L. T. Kohn, J. Corrigan, M. S. Donaldson, et al., To err is human: building a safer health system, vol. 6. National academy press Washington, DC, 2000.

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

[5] P. A. Sandercock, M. Niewada, and A. Członkowska, "The international stroke trial database," Trials, vol. 13, no. 1, pp. 1–1, 2012.

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

[7] S. Y. Adam, A. Yousif, and M. B. Bashir, "Classification of ischemic stroke using machine learning algorithms," Int J Comput Appl, vol. 149, no. 10, pp. 26–31, 2016.

[8] A. Sudha, P. Gayathri, and N. Jaisankar, "Effective analysis and predictive model of stroke disease using classification methods," International Journal of Computer Applications, vol. 43, no. 14, pp. 26– 31, 2012.

[9] G. Kaur and A. Chhabra, "Improved j48 classification algorithm for the prediction of diabetes," 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 Kaufmann, 2016.

[11] P. Sewaiwar and K. K. Verma, "Comparative study of various decision tree classification algorithm using weka," International Journal of Emerging Research in Management &Technology, vol. 4, pp. 2278– 9359, 2015.