Medical Based Decision Support System For Stroke Prediction And Its Types Using ML Algorithms

Jayashree KY, Divya AK, Prajna MR, Bhavya PS
College of Engineering, Mysuru

Abstract- *In the modern-day world, when compared with the other medical issues, Stroke is ranked second among the long-term & serious reasons for death. Stroke causes demise of cerebrum cells because acute supply of oxygen & bloodstream blockage. WHO has issued a warning on the possibility of increase in deaths due to stroke in the years to come. Although numerous advancements have taken place in the field of stroke to detect stroke in their early stages using AI & ML approach caused by different types of stroke attacks with the help of various forecasting methods to detect different types of strokes in humans. The work emphases improved& various stroke detecting techniques.*

Keywords- Machine Learning, AI, Data Science, Stroke, Types, KNN, Naïve Bayes, Binning Method.

I. INTRODUCTION

When compared with the other medical issues, Stroke is ranked second among the long-term & serious reasons for death. Stroke causes demise of cerebrum cells because acute supply of oxygen & bloodstream blockage. WHO has issued a warning on the possibility of increase in deaths due to stroke in the years to come. Although numerous advancements have taken place in the field of stroke to detect stroke in their early stages using AI & ML approach caused by different types of stroke attacks with the help of various forecasting methods to detect different types of strokes in humans. The work emphases improved & various stroke detecting techniques.

The below said data-set have been collected using various sources such as www.kaagle.com,www.data.world.com, www.data.gov.in. etc. Although numerous advancements have taken place in the field of stroke to detect stroke in their early stages using AI & ML approach caused by different types of stroke attacks with the help of various forecasting methods to detect different types of strokes in humans. The work emphases improved & various stroke detecting techniques.

Prediction of stroke is considered as a parameter for concern and worldwide health issue especially with the implementation technique being manual and consumes more time. The system proposed through this work helps predict stroke and its types during early stages using the techniques involving AI&ML. Predicting stroke infection and illness has been made possible with high reliability using automated method involving advancements of AI&ML techniques.

Rupture & coagulation of veins & blood compartments has been gradually increasing as per WHO records & will increase in the years to come resulting in loss of vision, motion and so on. On a constant basis, approximately 10 lakh individuals meet with stroke with the common symptom being vision and motion loss with discharge which in turn damage cerebrum & heart disrupting the circulation of blood and oxygen to the brain.

II. RELATED WORKS

1. Title: Automated Ischemic Stroke Sub typing Based on Machine Learning Approach

Authors: GANG FANG,PENG XUANDWENBIN LIU
Year of Publications: 2020

Description: Ischemic stroke subtyping being extremely valued for treatment &effective intervention & also for the forecast of ischemic stroke, with manual interpretation resulting in error and excessive consumption of time. This study uses an integrated MLtactic to categorize the ischemic stroke subtypes based on the dataset of The International Stroke Trial. Using Shapiro-Wilk algorithm importance of features were ranked and analysis were done using Pearson correlations. The result sexhibited that the approach in outclassed human specialists.

Methodology: Multinomial-Naïve-Bayes-Classier linear SVC, Random-Forest-Classier, Extra-Trees-Classier & AdaBoost-Classier.

LIMITS:

- Using techniques such as Rapid Miner,etc. the outcome can be effortlessly attained & the challenging is not probable.
- Data sets used are small for prediction.
- Fewer precise results.

2. Title:The Use of Deep Learning to Predict Stroke Patient Mortality

Authors: Songhee Cheon,JKimand Jihye Lim
Year of Publications: 2021

Description: Aging of Korean population with rapid stroke incidence will enforce a financial burden on the society. Investigation shave emphasized & implemented excerpt medical records with medical features to predict stroke. Comparison on have been made with other five machine-learning methods.

Methodology: Scaled PCA/deep neural network [DNN] approach**.**

LIMITATIONS:

- Only Model build.
- Not suitable for real time.
- Less accurate results.

3.Detection of Brain Stroke using Electroencephalography (EEG)

Authors: Pinanshu Garg, Prateek K, Kshitij Shakya, D Khurana Shubhajit R.

Year of Publications: 2020

Description: Interruption in life bloodstream causes Brain stroke with oxygen being poor which makes an individual to collapse. To identify brain stroke in a specific region normally we go with MRI & CT scan. With CT scan/ MRI, being expensive, extensive research has been done in this work to find a substitute method for the same with the EEG/MRI/CT data being used from the past 48 hours of patients being used.
Methodology: Image processing methods used**.**

LIMITATIONS:

- Image datasets used.
- Not suitable for real time.
- One type of stroke predicted.

4. Enabling Efficient Stroke Prediction by Exploring Sleep Related Features

Authors: Jia Xie, Zhu Wang, Zhiwen Yu, Bin Guo

Year of Publications: 2022

Description: Stroke is triggered by the progressive ailment of the chief nerve system and results in a grave influence further resulting worthy to allow primary analysis or forecast of stroke by nursing people's daily physical statistics and then scheming a valuable stroke predictor whenever the indications are not seeming. An ewtactic for stroke prediction by discovering sleep related features. Evaluating the outline using an actual polysomnogram data set consisting of 159 healthy individuals and 66 patients. Experimental results demonstrating that the planned model can fore cast stroke events much effectively, with Precision.

Methodology: data mining approach**.**

LIMITATIONS:

- Only Model built.
- Not suitable for real time.
- Less Datasets used.

5.Machine Learning Algorithm for Stroke Disease Classification

Authors: T Badriyah, N Sakinah, Iwan Syarif

Year of Publications: 2020

Description: Eight machine learning algorithms are used in this study namely Logistic Regression,K-Nearest Neighbors, etc. at(95.97%),(94.39%)& (96.12%) respectively.

Methodology: Machine learning techniques.

LIMITATIONS:

- Only Model build.
- Not suitable for real time.

2.1 Survey Summary

1. Prediction of IschemicStroke has been stated in the IEEE paper. But in the proposed system we predict all 3types using various techniques of Machine learning.
2. An idea of stroke prediction has been presented in the IEEE paper, Nevertheless in the proposed system depicting as a scenario which is being valuable for hospitals, doctors & patients.
3. Algorithm generating less accurate results and less parameters have been used for the prediction, with our projected system using effectual algorithms and 20 parameters being used.

4. Bayesian algorithm, "Decision Tree algorithm" have been implemented to procure good results.
5. Less amount of training data-sets with the proposed system having huge data-sets for processing are being utilized.
6. Algorithms used are not programmed in IEEE papers. But in the proposed system a unique reason for the algorithm has been written and outcomes will be verified.
7. Skills such as "*visual Studio "SQL Server*" and C language have been used for programming.

III. PROPOSED SYSTEM

Proposed system is considered as an automation to forecast. It is a browser-based application which can accessed from anywhere. Proposed system uses data preprocessing technique to remove irrelevant data and extract relevant data. System uses around 7000 records as training data-sets.Proposed system is beneficial in prediction at early stages.

IV. METHODOLOGY

Step 1:Collection of Raw data

Collection of Previous years data for processing have been used, comprises of information and constraints that are essential for forecast.

Step 2: Extraction and Segmentation of Data (Data Preprocessing)

Medical data are analyzed with extraction of only relevant data. Essential data abstraction is done as the complete training data is not essential for the processing.

Step 3: Data Training

After extraction and segmentation of data, training the data becomes essential changing the data into the obligatory set-up is done. Adaptation depends on the nature of algorithm.

Step 4: Prediction of disease using the technique of ML

Machine Learning

Machine learning algorithms are made use of to process data& study the system based on the available data.

Technique using supervised learning

Using a predictive model for the forecast of one value. Controlled learning includes labels which are predefined as they classify an object based on the constraints to one of the set of labels which are predefined.

As numerous algorithms are available to shape model in controlled learning such as KNN, Naive bayes, etc.

Bayesian Classifier or KNN algorithm is used for the prediction.

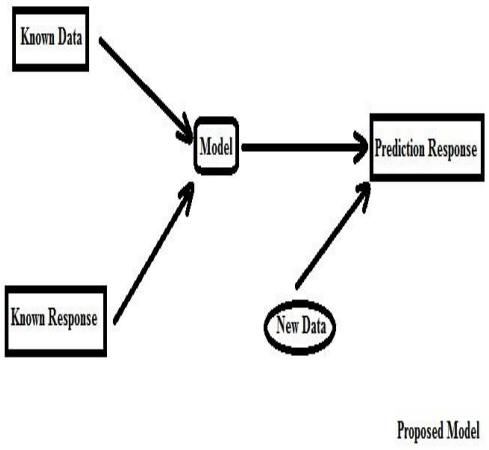

Fig: 1 Methodology

Step 5: Prediction of Stroke

Doctors can easily study to the module system predicting types of strokes for an ovel patient grounded on the input constraints.

Step 6: Outcomes

Algorithm is checked for accuracy using the method of confusion matrix.

Step 7: Visual Illustration

The obtained results are represented on GUI. Prediction and display of the disease on a GUI as the user logins to the system.

V. ALGORITHMS

5.1 KNN Algorithm Steps

Working of KNN ?

1.No of nearest neighbors for K are determined.

2. Distance calculation is performed using Euclidean, Manhattan.
3.K-minimum distance neighbors are determined in the process.
4.Congregationvalues.
5. Usingmeekmainstream to worth of qurey case is predicted.

5.2 KNN Algorithm Flow

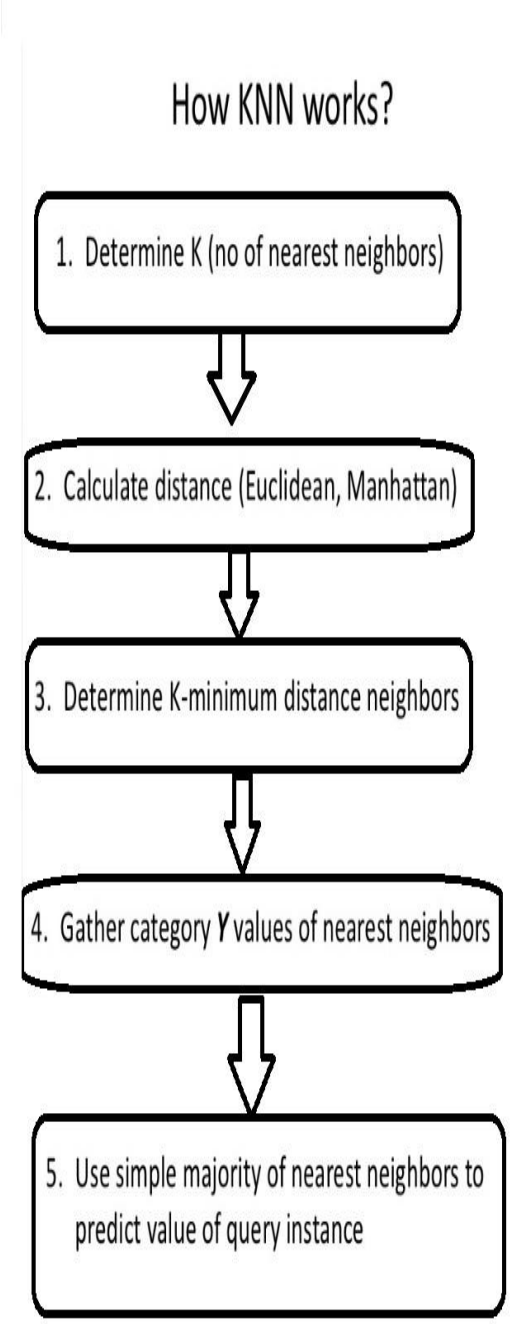

Fig 2: "KNN Algorithm Flow"

5.3 KNN Algorithm Pseudocode

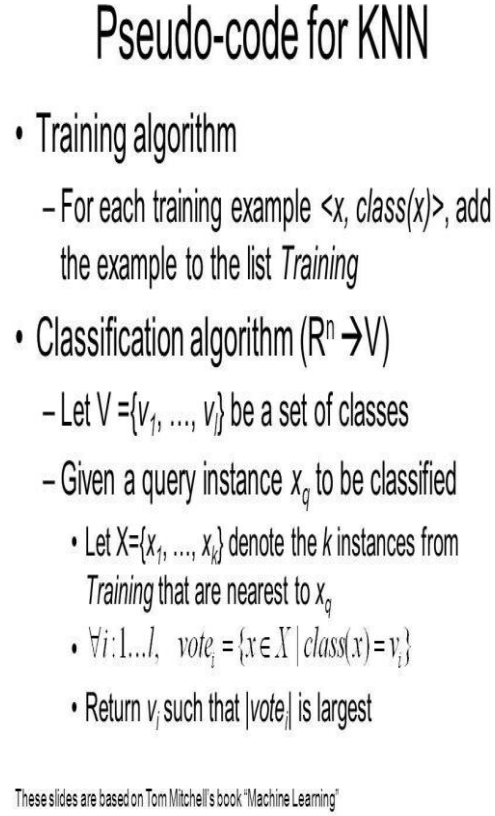

Fig 3: KNN Algorithm pseudocode

5.4 Naïve Bayes Algorithm

5.4.1 Phases

Phase 1: Scanning servers with data.

Processing of the Retrieved &vital data from the source.

Phase 2: Likeli hood calculation of each attribute value. [n, n_c, m, p]

calculate the probability of incidence using the formula for each trait is performed, applying the formula for each label.

Phase 3: Formula application

P(attribute value(ai)/subject value vj)=(n_c+mp)/(n+m)

"Where:"

- n=the no. of training examples forwhich v = vj
- nc = number of examples for which v = vj and a = ai
- p = a priori estimate for P(aijvj)
- m=the equivalent sample size

Phase 4: Probability multiplication

Multiplication of the trait results for each class using final results for organization.

Phase 5: Value comparison and classification of the trait values to one of the "predefined sets of class.

5.5 Navie Bayes Algorithm Flow

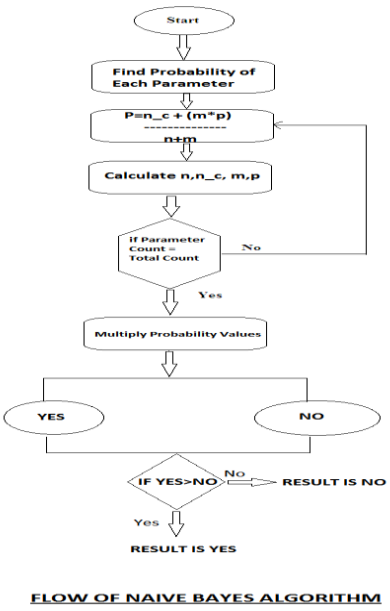

Fig: 4 – NB Algorithm Flow Diagram

5.6 Naïve Bayes Algorithm Pseudo-code

Input:

Training dataset T,

F= (f_1, f_2, f_3,.., f_n) // value of the predictor variable in testing dataset.

Output:

A class of testing dataset.

Step:

1. Read the training dataset T;
2. Calculate the mean and standard deviation of the predictor variables in each class;
3. Repeat

   Calculate the probability of f_i using the gauss density equation in each class;

   Until the probability of all predictor variables (f_1, f_2, f_3,.., f_n) has been calculated.
4. Calculate the likelihood for each class;
5. Get the greatest likelihood;

Pseudocode of naive bayes algorithm

Fig: 5 – NB Algorithm Pseudo-code

VI. RESULTS

6.1 Naive Bayes Algorithm Results

Constraint	Naive Bayes Algorithm
Accuracy	96.6 %
Time limit (M.sec)	2223
Appropriatelyorganized(accuracy)	96.6 %
Inappropriatelyorganized (Recall)	3.4 %

6.2 KNN Results

Constraint	KNN Algorithm
Accuracy	95%
Time limit (M.sec)	1780
Appropriately organized(accuracy)	95%
Inappropriately organized (Recall)	5%

VII. CONCLUSION

Being hard to forecast stroke in their initial phases, the projected structure helps clinicians to take quicker and healthier choices. The System practices AI techniques further add effectual and precise results. The most actual classifiers being "K nearest neighbor" and "Random Forest".

REFERENCES

[1] Duen-Yian Yeh a, Ching-Hsue Cheng b, Yen-Wen Chen b A predictive model for cerebrovascular disease using data mining "Science, Vol. 8970- 8977, 2011.

[2] Cheng-Ding Chang a, Chien-Chih Wang b, Bernard C. Jiang Using data mining techniques for multi-diseases prediction modeling of hypertension and hyperlipidemia by common risk factors Vol 38 ,5507–5513, 2011.

[3] Genetics and Genomics of Stroke Novel Approaches Alison E. Baird, MBBS, PHD Brooklyn, New York Vol. 56, No. 4, 2010.

[4] M. Anbarasi et. al. Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm International Journal of Engineering Science and Technology Vol. 2(10), 5370-5376 ,2010.

[5] Shanta Kumar B.Patil,Y.S.Kumaraswamy. „Predictive data mining for medical diagnosis of heart disease

prediction‟ JyotiSoni, ijmaAnsari, Dipesh Sharma IJCSE Vol .17, 2011 .
[6] https://www.researchgate.net/publication/5880405