International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Detection of Stroke Disease Using Machine Learning

Kavyashree CC[1], Srividya A[2], Pavithra S[3], Mohammed Salamath[4], Priyanka M N[5]

[1]ISE, MITT, Mysuru, India
kavyashreecc3[at]gmail.com

[2]ISE, MITT, Mysuru, India
asrividya01[at]gmail.com

[3]ISE, MITT, Mysuru, India
pavithra052002[at]gmail.com

[4]Professor, ISE, MITT, Mysuru, India
salamath1920[at]gmail.com

[5]ISE, MITT, Mysuru, India
priyankamn15[at]gmail.com

Abstract: ***Stroke Type prediction has become a global health issue nowadays and is an area of concern. Current system is a manual, time consuming, requires more experience of doctors and expensive and leads to less accurate results as it is manual decisions. Stroke prediction is one of the increasing diseases in the current medical sector. Prediction of Stroke disease is difficult at early stages our proposed system helps to predict the stroke disease and related types at early stages using ML algorithms. Most investigations performed on the robotized analysis of stroke and its sub - types were on the picture preparing methods and CT scan and MRI. An artificial neural system [7] gives a general method for moving toward issues. An Artificial neural system - based expectation of stroke illness enhances the analytic exactness with higher consistency.***

Keywords: Stroke using Machine Learning, Machine Learning algorithms, K Nearest Neighbors, and Random Forest, 3types of stroke Ischemic stroke, Hemorrhagic stroke, Transient Ischemic stroke

1. Introduction

Stroke is a second leading disease which causes death and they have been serious. Stroke is the sudden loss of cerebrum cells because of less intake of oxygen, caused by blockage of bloodstream or break of a supply route to the mind. According to World Health Organization in the upcoming years this disease will increase the death rate. Stroke is a mind assault and it takes place at any time. It causes coagulation in veins or break of blood compartments. As indicated by world health organization stroke will keep on increasing in coming years, so earnest condition treatment must be very quick. It is found that a million individuals overall endure a stroke. Causes of this disease are loss of vision, loss of motion, and discourse. Stroke is a significant ailment which hurts the cerebrum [1], like [5] heart strike which hurts the heart. It doesn't circulate blood and enough oxygen to the brain cells. Everyone overcomes with some stroke peril. This disease may cause loss of movement, sudden torment in the chest, talk inability, loss of memory and thinking limit, dazelike state, or passing. This impacts the person of all ages. It can be balanced through riskfactors [2] which are essential. It is said that the most surely understood remedial error happens in light of expiry of medicines misguided estimations and treatment given to the wrong patient.

Stroke composes of three types:

- Ischemic stroke
- Hemorrhagic stroke
- Transient ischemic stroke

2. Related Work

1) Enabling efficient stroke prediction by exploring sleep related features

Year of Publication: 2018

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

Methodology: They have used readily available data tools such as Rapid Miner tool for implementation.

Limitations:

- Using tools like Rapid Miner and other such tools namely RTool, Wekaa tool the results can be easily obtained but the testing of these is not possible.
- Small Data - set used for prediction.
- Less accurate results.

2) Comparison of different machine learning approaches to model stroke subtype

Classification and risk prediction.

Year of publication: 2018

Authors: Luis Garcia - Temza, Jose L. Risco - Martin, Jose L. Ayala, Gemma Reig Rosello, Juan M. Camarasaltas.

Methodology: Random Forest and support vector machine (SVM) is used for implementation.

Limitations:

- These Random Forest and SVM techniques produce graphical outputs the distinguishing will be difficult in the graphical method.
- Very Small Data - sets used.

3) Prediction and Control of Stroke by Data Mining.
Year of publication: 2016
Authors: Fariborz Khorvash, Rasul Norouzi.
Data set used: The data set used for implementation is only static data available on the UCI Machine Learning Repository.
Data Mining tools used.

Limitations:

- Uses data mining techniques.
- Huge data required.
- More time required for prediction.

4) An Artificial Intelligence Approach for Predicting Different Types of Stroke.
Year of Publication: 2018
Authors: Hima Haridas Dept. of Computer science Jyothi Engineering College Thrissur, India.
Aswathy Wilson Dept. of Computer science Jyothi Engineering College Thrissur, India.
Methodology: Most investigations performed on the robotized analysis of stroke and its subtypes were on the picture preparing methods and CT scan and MRI.

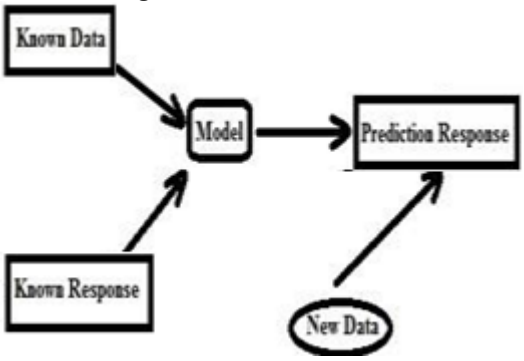
Limitations: Detects only after the stroke has occurred.

5) Machine Learning Algorithm for Stroke Disease Classification.
Year of publication: 2020
Authors: Tessy Badriyah Politeknik Elektronika Negeri, Surabaya, INDONESIA tessy[at]pens.ac.id
Nur Sakinah Politeknik Elektronika Negeri
Surabaya, INDONESIA, nursakinah[at]pasca. student. pens. ac. id
Daisy Rahmania Syarif, University of Cologne, GERMANY, Iwan Syarif, dsyarif[at]smail. unikoeln. De Politeknik Elektronika Negeri Surabaya, Indonesia, iwanarif[at]pens. ac. id
Methodology: System Design is a system design used in this study, which consists of 3 (three) main stages: Data Collection, Pre - processing data and Performance analysis method of classification.

Limitations: It is time consuming, and not accurate.

Machine Learning: Machine learning is a process of studying a system based on data. Machine learning is a part of data science where we use machine learning algorithms to process data.

Supervised Learning Technique: It's a predictive model used for the tasks where it involves prediction of one value using other values in the data - set. Supervised learning will have predefined labels. It classifies an object based on the parameters to one of the predefined set of labels. We have many algorithms to build model in supervised learning such as KNN, Naive bayes, Decision Tree, ID3, Random Forest, SVM, Regression techniques etc. Depending of the requirement, labels, parameters and data - set we select the appropriate algorithm for predictions. Algorithm is used to build a model that makes predictions based on evidence in the presence of uncertainty. In this project for prediction, we make use to "*Bayesian Classifier or KNN algorithm*" which is an efficient and works fine for all different sets of parameters. It also generates accurate results.



Random Forest Algorithm Steps

Operation of Random Forest
The working of Random Forest Algorithm as follows.

1) A random seed is chosen which pulls out at random collection of samples from training dataset while maintaining the class distribution
2) With this selected data set, a random se of attributes from the original data set is chosen based on user defined values. All the input variables are nt considered because of enormous computation and high chance of overfitting
3) In dataset where M is the total number of input attributes in the dataset, only R attributes are chosen at random for each tree where R<M.
4) The attributes from this set creates the best possible split using the gini index to develop a decision tree model. The process repeats for each of the branches until the termination condition stating that leaves are the nodes that are too small to split

Bayesian Classifier Algorithm Steps:

Step 1: Scan the dataset (storage servers) retrieval of required data for mining from the servers such as database, cloud, excel sheet etc.

Step 2: Calculate the probability of each attribute value. [n, n_c,
Here for each attribute we calculate the probability of m, p] occurrence using the following formula (mentioned in the next step). For each class (disease) we should apply the formulae.

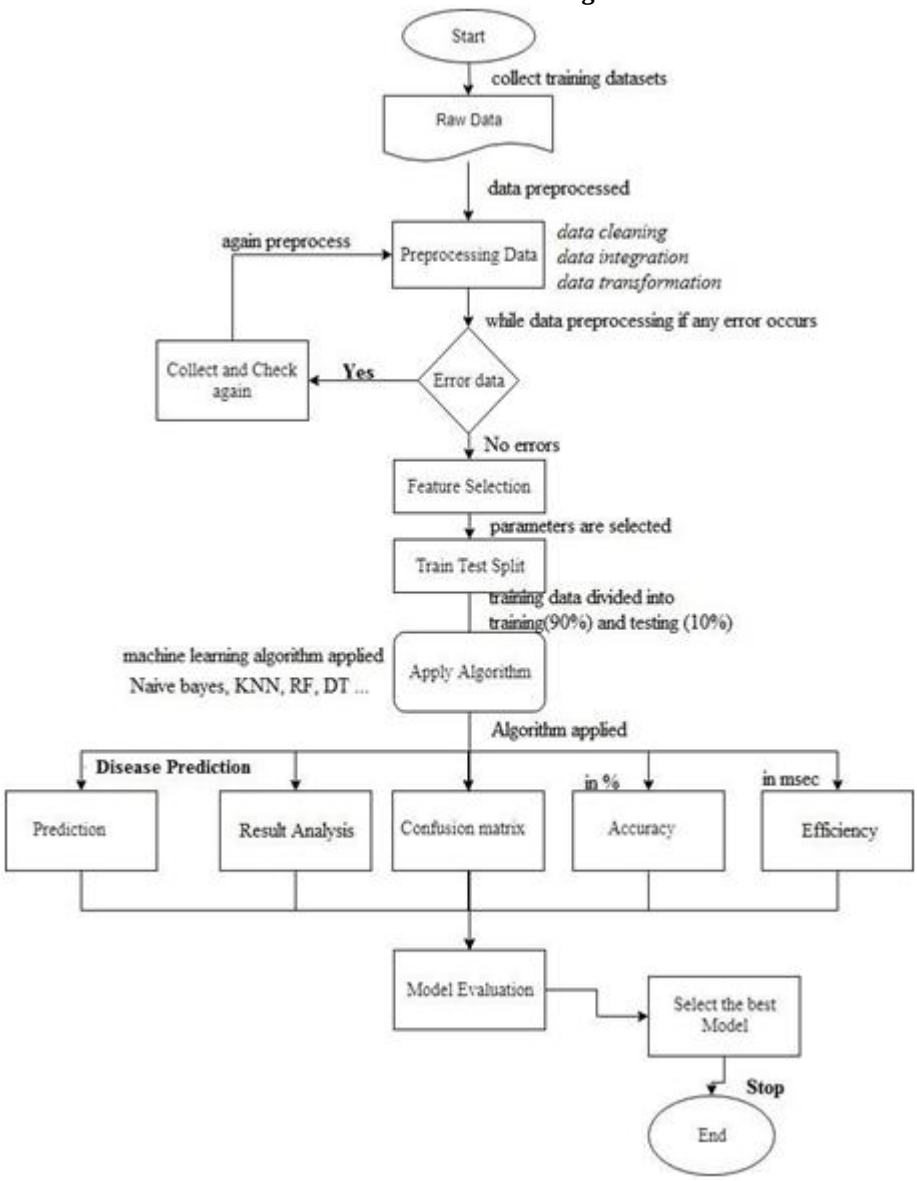
Step 3: Apply the formulae
P (attribute value (ai)/ subject value vj)=(n_c + mp)/(n+m)
Where
n = the number of training examples for which v = vj
nc = number of examples for which v = vj and a = ai p = a priori estimate for P (aijvj) m = the equivalent sample size

Step 4: Multiply the probabilities by p

For each class, here we multiple the results of each attribute with p and final results are used for classification.

Step 5: Compare the values and classify the attribute values to one of the predefined set of class.

Flow Diagram



3. Conclusion

Stroke prediction is one of the leading diseases in the current medical sector. As it is difficult to predict the stroke disease andit's types at early stages, this proposed system helps to predict the stroke disease and related types at early stages. Our system is useful for doctors to take faster and better decisions. System uses ML techniques to get efficient and accurate results. We have used most efficient classifiers such as "K NearestNeighbor" algorithm, and "Random Forest" algorithm. System is a real time application useful for medical sector.

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. JiangUsing 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 AlisoBaird, 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), 53705376, 2010.

[5] ShantakumarB. Patil, Y. S. Kumaraswamy, Predictive data mining for medical diagnosis of heart disease prediction "jyotisoni, ujmaansari, dipeshsharma IJCSE Vol.17, 2011.

[6] https: //www.researchgate. net/publication/5880405

[7] Lela Mirtskhulava, Gillian Pearce et al Artificial Neural Network Model in Stroke Diagnosis 2015 17th

UKSIM - AMSS International Conference on Modelling and Simulation.

[8] Mirtskhulava Lela, Al - Majeed Salah, Gillian Pearce, Gogoladze Tamar, Ivane Javakhishvili. Blood clotting prediction model using Artificial Neural Networks and Sensor Networks. GESJ: Computer Science and Telecommunications 2014|No.3 (43). ISSN 1512 - 1232. Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018) IEEE Xplore Compliant - Part Number: CFP18BAC - ART; ISBN: 978 - 1 - 5386 - 1974 - 2 978