International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 10, ISSN No. 2455-2143, Pages 72-75
Published Online February 2023 in IJEAST (http://www.ijeast.com)

ML APPROACH FOR BRAIN STROKE PREDICTION USING IST DATABASE

A.P.V Rohit, M.Umesh Chowdary, G.B.S Ashish, V.Anitha
Department of CSSE
Lendi Institute of Engineering & Technology

Swaroop Sana
Assistant Professor, Department of CSSE,
Lendi Institute of Engineering & Technology

Abstract— **Stroke is one of the most serious diseases worldwide, directly or indirectly responsible for a significant number of deaths. Various data mining techniques are used in the healthcare industry to aid in the diagnosis and early detection of diseases. Current research considers several elements that lead to stroke. First, we examine the characteristics of those who suffer a stroke more often than others. The dataset is from a freely available source and various classification algorithms are used to predict the onset of a stroke shortly. Using the Naïve Bays and Decision Tree, it was possible to achieve an accurate percent. Using various statistical techniques and principal component analysis, we identify the most important factors for stroke prediction. We conclude that age, heart disease, average glucose level, and hypertension are the most important factors for detecting stroke in patients by using Machine Learning Approach.**

Keywords— **Machine Learning, Stroke, Classification, Supervised Learning, NB**

I. INTRODUCTION

Stroke is a ailment that impacts vessels that supply blood to the thoughts. mind stroke takes region which list blood glide to the mind is each reduced or interrupted. whilst this occurs, the mind no longer gets sufficient oxygen or other crucial components, and the brain cells start to die. A stroke effects important lengthy-time period incapacity or demise. mind stroke is one of the leading causes of death all around the world. There are 3 kinds of brain strokes: ischemic strokes, hemorrhagic strokes, and transient ischemic assault (TIA), which is also referred to as a caution or mini-stroke. Ischemic strokes arise due to loss of blood supply, and hemorrhagic strokes occur because of ruptured blood vessels.

The most typical kind IS ISCHEMIC STROKE is this one. It occurs when the blood arteries in the brain narrow or block, significantly reducing the amount of blood flow (ischemia). Fat deposits that accumulate in blood vessels or blood clots or other debris that move through the bloodstream, typically from the heart, and lodge in the blood vessels in the brain cause blocked or restricted blood arteries.

Brain bleeding results in a hemorrhagic stroke. This may occur when a brain blood artery rupture or when bleeding occurs in the brain tissue. Pressure brought on by bleeding, edema, or a lack of blood flow can all contribute to hemorrhagic stroke damage. An ischemic stroke, which is a stroke brought on by a stopped blood supply, can result in bleeding in the brain tissue. As a result, the brain's tissue is harmed.

Transient ischemic attack, or TIA for short, is a dangerous repercussion. A TIA causes a temporary interruption in the blood supply to a portion of the brain. Another name for it is a "ministroke," but don't be deceived by the diminutive. A TIA may be a precursor to a full- blown stroke. The most common cause of TIAs is a blood clot that becomes stuck in an artery that carries blood to the brain. Your brain is oxygen-starved and unable to function normally if there isn't regular blood flow

Machine learning is a branch of Artificial Intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, uncovering key insights within data mining projects. These insights subsequently drive decision-making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase, requiring them to assist in the identification of the most relevant business questions and subsequently the data to answer them.

Machine Learning classifiers are used for different purposes and these can also be used for detecting fake news. Machine learning methods enable computers to operate autonomously without explicit programming. ML applications are fed with new data, and they can independently learn, grow, develop, and adapt. Machine learning works on data and it will learn through some data. Machine learning is very different from

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 10, ISSN No. 2455-2143, Pages 72-75
Published Online February 2023 in IJEAST (http://www.ijeast.com)

the traditional approach. In, Machine learning we fed the data, and the machine generates the algorithm. Machine learning has three types of learning

- Supervised learning
- Unsupervised learning
- Reinforcement learning

II. LITERATURE SURVEY

Stroke Prediction Using SVM

In this paper we were using Support Vector Machine for stroke prediction. 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). Machine learning algorithms have been proposed as important tools indecision making in medical field. The objective of this work is to develop a machine learning based approach to predict the possibility of stroke in people having the symptoms or risk factors of stroke. In this work, we have implemented SVM with different kernel functions and found that linear kernel gave an accuracy of 90 %. The results were evaluated on a spectrum of patients of different age groups.

Stroke Prediction using Artificial Intelligence

The stroke deprives person's brain of oxygen and nutrients, which can cause brain cells to die. Numerous works have been carried out for predicting various diseases by comparing the performance of predictive data mining technologies. In this work, we compare different methods with our approach for stroke prediction on the Cardiovascular Health Study (CHS) dataset. Here, decision tree algorithm is used for feature selection process, principle component analysis algorithm is used for reducing the dimension and adopted back propagation neural network classification algorithm, to construct a classification model. The proposed method use Decision Tree algorithm for feature selection method, PCA for dimension reduction and ANN for the classification. The experimental results show that the proposed method has higher performance than other related well-known methods.

Burden of Stroke in the World

Stroke is the second leading cause of death and leading cause of adult disability worldwide with 400-800 strokes per 100,000, 15 million new acute strokes every year, 28,500,000 disability adjusted life-years and 28-30-day case fatality ranging from 17% to 35%. The burden of stroke will likely worsen with stroke and heart disease related deaths projected to increase to five million in 2020, compared to three million in 1998. This will be a result of continuing health and demographic transition resulting in increase in vascular disease risk factors and population of the elderly. Developing countries account for 85% of the global deaths from stroke. The social and economic consequences of stroke are substantial. The cost of stroke for the year 2002 was estimated to be as high as $49.4 billion in the United States of America (USA), while costs after discharge were estimated to amount to 2.9 billion Euros in France.

Burden of Stroke in Uganda

The actual burden of stroke in Uganda is not known. According to WHO estimates for heart disease and stroke 2002, stroke was responsible for 11 per 1000 population (25,004,000) 4 disability adjusted life years and mortality of 11,043. Stroke is one of the common neurological diseases among patients admitted to the neurology ward at Mulago, Uganda's national referral hospital accounting for 21% of all neurological admissions. Unpublished research done at Mulago hospital, showed a 30-day case fatality of43.8% among 133 patients admitted with stroke. The economic burden caused by stroke has not been explored in Uganda but given the very high dependent population (53%), high prevalence of HIV/AIDS, drug resistant TB and Malaria, the impact of stroke and other emerging non-communicable diseases on the resource limited economy is astronomical.

AI-Based Stroke Disease Prediction System Using Real-Time Electromyography Signals

In this paper, we developed a stroke prediction system that detects stroke using real- time bio- signals with artificial intelligence (AI). Both machine learning (Random Forest) and deep learning (Long Short-Term Memory) algorithms were used in our system. EMG (Electromyography) bio-signals were collected in real time from thighs and calves, after which the important features were extracted, and prediction models were developed based on everyday activities. Prediction accuracies of 90.38% for Random Forest and of 98.958% for LSTM were obtained for our proposed system. This system can be considered an alternative, low-cost, real-time diagnosis system that can obtain accurate stroke prediction and can potentially be used for other diseases such as heart disease.

III. PROPOSED ALGORITHM

Now, to overcome the drawbacks of SVM in our project to deal with these problems by introducing the Naïve Bayes and Decision Tree. This system can predict the disease but not the sub type of the disease and it fails to identify the condition of the people. To deal with these problems by introducing these algorithms to get accurate results. The SVM is not given accurate result, to resolved that by using the Naive Bayes and Decision Tree.

IV. MACHINE LEARNING METHODOLOGY

Using this methodology, the modeler can discover the "performance ceiling" for the data set beforesettling on a model. In many cases, a range of models will be equivalent in terms of performance so the practitioner can weigh the benefits of different methodologies. Few methodologies used in our projects are:

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 10, ISSN No. 2455-2143, Pages 72-75
Published Online February 2023 in IJEAST (http://www.ijeast.com)

- Decision Tree
- Naïve Bayes

Decision Tree

A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements. Decision tree is one of the important methods for handling high dimensional data. Tree based learning algorithms are considered to be one of the best and mostly used supervised learning methods.

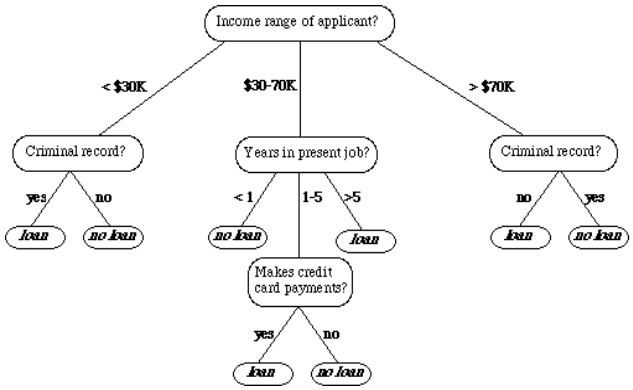


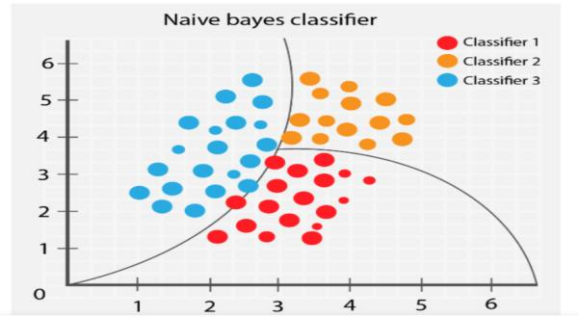
Fig. 1. Decision Tree Algorithm

Naïve Bayes

A Naïve Bayes classifier is a probabilistic machine learning model that's used for classification task. The crux of the classifier is based on the Bayes theorem. Using Bayes theorem,

P(A|B) = P(B|A) * P(A) / P(B)

we can find the probability of A happening, given that B has occurred. Hence, B is the evidence and A is the hypothesis. The assumption made here is that the predictors/features are independent. That is the presence of one particular feature does not affect the other. Hence it is called naïve.



V. MATERIALS & METHODS

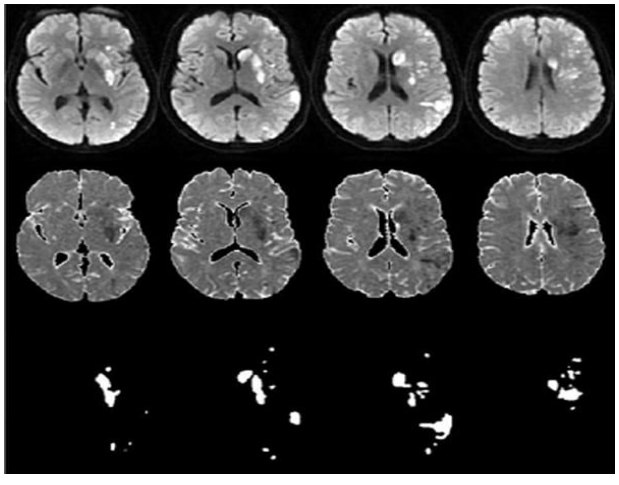
3.1. Dataset Description

Our research was based on a dataset from Kaggle [34]. From this dataset, we focused on participants who are over 18 years old. The number of participants was 3254, and all of the attributes (10 as input to ML models and 1 for target class) are described as follows:

• Age (years) [39]: This feature refers to the age of the participants who are over 18 years old.
• Gender [39]: This feature refers to the participant's gender. The number of men is 1260, whereas the number of women is 1994.
• Hypertension [40]: This feature refers to whether this participant is hypertensive or not. The percentage of participants who have hypertension is 12.54%.
• Heart_disease [41]: This feature refers to whether this participant suffers from heart disease or not. The percentage of participants suffering from heart disease is 6.33%.
• Ever married [42]: This feature represents the marital status of the participants, 79.84% of whom are married.
• Work type [43]: This feature represents the participant's work status and has 4 categories (private 65.02%, self-employed 19.21%, govt_job 15.67% and never_worked 0.1%).
• Residence type [44]: This feature represents the participant's living status and has 2 categories (urban 51.14%, rural 48.86%).
• Avg glucose level (mg/dL) [45]: This feature captures the participant's average glucose level.
• BMI (Kg/m2) [46]: This feature captures the body mass index of the participants.

- Smoking Status [47]: This feature captures the participant's smoking status and has 3 categories (smoke 22.37%, never smoked 52.64% and formerly smoked 24.99%).

• Stroke: This feature represents if the participant previously had a stroke or not. The percentage of participants who have suffered a stroke is 5.53%.



VI. REFERENCE

[1] Learn about Stroke. Available online: https://www.world-stroke.org/world-stroke-day-campaign/why-stroke-matters/learnabout-stroke (accessed on 25 May 2022).

[2] Elloker, T.; Rhoda, A.J. The relationship between social support and participation in stroke: A systematic review. Afr. J. Disabil. 2018, 7, 1–9. [CrossRef] [PubMed]

[3] Katan, M.; Luft, A. Global burden of stroke. In Seminars in Neurology; Thieme Medical Publishers: New York, NY, USA, 2018; Volume 38, pp. 208–211.

[4] 4. Bustamante, A.; Penalba, A.; Orset, C.; Azurmendi, L.; Llombart, V.; Simats, A.; Pecharroman, E.; Ventura, O.; Ribó, M.; Vivien, D.; et al. Blood biomarkers to differentiate ischemic and hemorrhagic strokes. Neurology 2021, 96, e1928–e1939. [CrossRef][PubMed]

[5] Xia, X.; Yue, W.; Chao, B.; Li, M.; Cao, L.; Wang, L.; Shen, Y.; Li, X. Prevalence and risk factors of stroke in the elderly in Northern China: Data from the National Stroke Screening Survey. J. Neurol. 2019, 266, 1449–1458. [CrossRef] [PubMed]

[6] Alloubani, A.; Saleh, A.; Abdelhafiz, I. Hypertension and diabetes mellitus as a predictive risk factors for stroke. Diabetes Metab. Syndr. Clin. Res. Rev. 2018, 12, 577–584. [CrossRef]

[7] Boehme, A.K.; Esenwa, C.; Elkind, M.S. Stroke risk factors, genetics, and prevention. Circ. Res. 2017, 120, 472–495. [CrossRef]

[8] Mosley, I.; Nicol, M.; Donnan, G.; Patrick, I.; Dewey, H. Stroke symptoms and the decision to call for an ambulance. Stroke 2007,38, 361–366. [CrossRef]

[9] Lecouturier, J.; Murtagh, M.J.; Thomson, R.G.; Ford, G.A.; White, M.; Eccles, M.; Rodgers, H. Response to symptoms of stroke in the UK: A systematic review. BMC Health Serv. Res. 2010, 10, 1–9. [CrossRef]

[10] Gibson, L.; Whiteley, W. The differential diagnosis of suspected stroke: A systematic review. J. R. Coll. Physicians Edinb. 2013, 43, 114–118. [CrossRef]

[11] Rudd, M.; Buck, D.; Ford, G.A.; Price, C.I. A systematic review of stroke recognition instruments in hospital and prehospital settings. Emerg. Med. J. 2016, 33, 818–822. [CrossRef]

[12] Delpont, B.; Blanc, C.; Osseby, G.; Hervieu-Bègue, M.; Giroud, M.; Béjot, Y. Pain after stroke: A review. Rev. Neurol. 2018, 174, 671–674. [CrossRef]

[13] Kumar, S.; Selim, M.H.; Caplan, L.R. Medical complications after stroke. Lancet Neurol. 2010, 9, 105–118. [CrossRef]

[14] Ramos-Lima, M.J.M.; Brasileiro, I.d.C.; Lima, T.L.d.; Braga-Neto, P. Quality of life after stroke: Impact of clinical and sociodemographic factors. Clinics 2018, 73, e418. [CrossRef]

[15] Gittler, M.; Davis, A.M. Guidelines for adult stroke rehabilitation and recovery. JAMA 2018, 319, 820–821. [CrossRef]

[16] Pandian, J.D.; Gall, S.L.; Kate, M.P.; Silva, G.S.; Akinyemi, R.O.; Ovbiagele, B.I.; Lavados, P.M.; Gandhi, D.B.; Thrift, A.G. Prevention of stroke: A global perspective. Lancet 2018, 392, 1269–1278. [CrossRef]

[17] Feigin, V.L.; Norrving, B.; George, M.G.; Foltz, J.L.; Roth, G.A.; Mensah, G.A. Prevention of stroke: A strategic global imperative. Nat. Rev. Neurol. 2016, 12, 501–512. [CrossRef]

[18] Fazakis, N.; Kocsis, O.; Dritsas, E.; Alexiou, S.; Fakotakis, N.; Moustakas, K. Machine learning tools for long-term type 2 diabetes risk prediction. IEEE Access 2021, 9, 103737–103757. [CrossRef]

[19] Alexiou, S.; Dritsas, E.; Kocsis, O.; Moustakas, K.; Fakotakis, N. An approach for Personalized Continuous Glucose Prediction with Regression Trees. In Proceedings of the 2021 6th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), Preveza, Greece, 24–26 September 2001; IEEE: Piscataway, NJ, USA, 2021; pp. 1–6.

[20] Dritsas, E.; Alexiou, S.; Konstantoulas, I.; Moustakas, K. Short-term Glucose Prediction based on Oral Glucose Tolerance Test Values. In Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies—HEALTHINF