



A Review of Predictive Systems for Patients at Risk of Developing Stroke

Oguntimilehin A¹, Olatunji K.A², Abiola O.B³, Akpan E.S⁴

^{1,2,3,4}Department of Computer Science, Afe Babalola University, Ado-Ekiti, Nigeria, ebenabiodun2@yahoo.com

Received Date : November 02, 2021 Accepted Date : November 24, 2021 Published Date : December 07, 2021

ABSTRACT

Stroke remains one of the leading causes of death worldwide. It is usually associated with a build-up of fatty deposits inside the arteries which increases the risk of blood clotting. The unannounced nature of the disease when it strikes has posed a major challenge in the health sector. Poor medical facilities, insufficient information on how to accurately diagnose stroke, late identification of the disease by the patients due to being ignorant of the disease are some of the reasons for the increasing mortality rate due to it. The application of data mining technique in the field of medicine has brought about positive development in the area of diagnosing, prediction and deeply understanding of healthcare data. This study considers some of the Predictive Models developed using some data mining approaches to predict patients at risk of developing stroke in order for other researchers to build on.

Key words : Stroke, Arteries, Data Mining, Predictive Models, Diagnosis

1. INTRODUCTION

The American Heart and Stroke Association defined stroke as a neurological disorder that happens when a brain cells die due to oxygen and nutrient deficiency. Previously, stroke was formerly called Cerebro-Vascular Accident (CVA) which recently is subdivided into three different types, which include; bleeding, Acute CVA, or Transient Ischemic Attack [10]. It is also discovered that occlusions of brain vas by a clot or vas rupturing are the main causes of oxygen and nutrient supply deficiency. Cardiovascular disease (CVD) is usually associated with a build-up of fatty deposits inside the arteries (atherosclerosis) and an increased risk of blood clots. It can also be associated with damage to arteries in organs such as the brain, heart, kidneys and eyes. CVD is not a single disease, but a cluster of diseases and injuries that affect the cardiovascular system (the heart and blood vessels). Stroke being a type of cardiovascular disease is defined as abrupt onset of a focal neurological deficit lasting more than 24 hours. It is also called cerebrovascular accident (CVA) or apoplexy [8].

A stroke is also called a “cerebral attack”, or even a “cerebral infarction”. In the event of a stroke, the symptoms persist and a brain injury is visible most often upon a radiological examination (scanner, magnetic resonance imaging). Some of the causes of strokes include: large blood vessel disease, small blood vessel disease, cardiac embolism, aortic or vertebral artery dissections, age and hereditary. Age and hereditary predisposition are the two major risk factors that one cannot change. According to the World Health Organization, the estimated global lifetime risk of stroke from the age of 25 years onward was 24.9% (95% confidence interval, 23.5 to 26.2); the risk among men was 24.7% (95% confidence interval, 23.3 to 26.0); and the risk among women was 25.1% (95% confidence interval, 23.7 to 26.5). The risk of Ischemic stroke was 18.3%, and the risk of hemorrhagic stroke was 8.2%. The mean global Burden of Disease 2016 Lifetime Risk of Stroke collaborators showed the mean global lifetime risk of stroke increased from 22.8% in 1990 to 24.9% in 2016, a relative increase of 8.9% (95% confidence interval, 6.2% - 11.5%); competing risk of death of any cause other than stroke. The 2020 reports of the World Health Organization on cardiovascular diseases shows they kill approximately 17.9 million people yearly [22].

Medical informatics is the study and application of methods to improve the management of patient data, clinical knowledge, population data, and other information relevant to patient care and community health [21]. The increased in demand for high-quality medical services coupled with the explosive growth of medical knowledge led to the suggestion that computer programs should be used in assisting physicians and other healthcare providers in discharging their clinical roles such as diagnosis, therapy and treatment[15] and such activities do employ data mining techniques among others.

2. REVIEW OF RELATED LITERATURES

To improve data analysis process in large and complex datasets, data mining researchers have long studied the application of tools and equipment. The introduction of data mining techniques in the field of medicine has brought about positive development in the area of diagnosing, predicting and deeply understanding of healthcare data. These

applications include treatment centers analysis aimed at improving treatment policies and prevention of any mistake in hospitals, early diagnosis of diseases, prevention of diseases and hospital death reduction [18].

A couple of researchers worked on stroke prediction, such as [23], reporting on heart disease diagnosis research using Asha and Sophia's 2010 [2] classification based on supervised machine learning. Study showed 10 fold cross validation was used to evaluate the data, and comparing the results. The experiment was performed using training data set consisting of 3000 instances with 14 different attributes. After the data were divided into two parts, three techniques were used for the testing set. The result showed their accuracy as follows: Naives Bayes (52.33%), Decision List (52%) and K-NN (45.67%).

Leila *et al.* [11], predicted the incidence of stroke using two different data mining techniques in the paper "prediction and control of stroke by data mining"; K-nearest Neighbour and C4.5 Decision Tree were used. Information on 807 healthy and sick subjects was collected, while 50 risk factors were used. After applying the algorithms and performing a comparison and computation of accuracy using the decision tree and K-Nearest Neighbour methods, the best results are those pertaining to the C4.5 algorithm that outruns the K-Nearest Neighbour algorithm in the Accuracy, Precision and Specificity criteria by a small difference. The results were: C4.5 Decision Tree (95.4%) and K-Nearest Neighbour (94.2%).

Moloud *et al.*, in [13] used and compared data mining techniques to predict the occurrence of heart disease using five different algorithms, C5.0, K – Nearest Neighbour, Support Vector Machine, Neural Network, and Logistic Regression, using 13 medical attributes with accuracy measures as 93.02%, 88.37%, 86.05%, 80.23% and 85.22% respectively, which shows that C5.0 gave the best performance during the prediction.

Shadman *et al.* [20] used five different machine learning algorithms to worked on a research “Heart Disease Prediction”. Dataset consisting of 566 records with 13 attributes we used. Support Vector Machine, Naïve Bayes, ANN, Random Forest and Simple Logistics were used on the datasets to build the model using 10 fold cross validation method on WEKA 3.8.2 application tool. The results of their accuracies showed that Support Vector Machine(SVM) provided the highest accuracy with 97.53%, followed by Random Forest with 95.76%. Others were Simple Logistics (95.05%), Naives Bayes (86.40%) and ANN (77.39%).

Boshra and Misaeid [5], in their study on Prediction and Diagnosis of Heart Disease by Data Mining Techniques, used WEKA tool which is a powerful data mining tool to apply the data mining algorithms. The experimental results from implementation of selected classification algorithms, J48

decision tree, Naive Bayes, K-Nearest Neighbour (K-NN) and Sequential Minimal Optimization (SMO) on heart disease dataset to analyze and compare the performance of the classifiers showed that the accuracy results were as follows: J48 Decision Tree (83.73), KNN (82.76), Naives Bayes (81.82) and SMO (82.76). After comparison, results showed that the best classification accuracy is 83.73% from J48 decision tree.

A clinical Diagnostic Model Based on Supervised Learning was presented by Ogutimilehin *et al.*[16]. The research motivations include: Shortages of medical practitioners, medical facilities and complexity of diseases among others gave rise to the need for the use of computer programs to give helping hands in the health sector. Method: The Partial Tree (PART) was used on the training instances to generate a classification model for the diagnosis.

Sabibullah and Kasmir [19] , presented Prediction of Stroke Risk through Stacked Topology of ANN Model. Research motivation include: Stroke being responsible for 3 million deaths in developing countries and also a major cause of deaths in Asian Countries. Estimation of stroke risks in population is not only helpful for healthcare providers but also important to identify persons at elevated risk and to select proper treatments in clinical trials. Method: The Stacked Topology Artificial Neural Network (ST-ANN) model was developed for the use of stroke risk prediction is trained, cross-validated and tested with the extracted features from three different stroke disease data sets. The accuracy rates (94%) achieved by the ST-ANN model presented for the risk prediction of stroke found to be higher than that of the stand-alone ANN model used in our earlier study (90.90% accuracy).

Improved study of Heart Disease Prediction System using data mining classification techniques was presented by Chaitrali and Sulabha [6]. The research was motivated by the challenge of quality of service to patients; poor diagnosis which has led to many disastrous consequences. Over 17millions total global deaths surveyed by World Health Organization (WHO) are due to heart attacks and stroke. In the research methodology, three data mining methods were applied namely Decision Tree, Naïve Bayes, and Neural Networks. The data set which consists of total 573 records in Heart Disease database was used of which 303 records was used as training set and 270 as testing set. The data mining tool WEKA 3.6.6 was used for the experiment. In the result, Neural Network provided accurate results as compared to Decision Tree and Naïve Bayes with results as follows: Neural Network (99.25% with 13 attributes and 100% with 15 attributes), Decision Tree (96.66% with 13 attributes and 99.62% with 15 attributes) and Naïve Bayes (94.44% with 13 attributes and 90.74% with 15 attributes).

3. SUMMARY OF THE FINDINGS FROM THE REVIEW

The review shows that a number of predictive systems or models are available today on the prediction and diagnosing of patients with stroke and those at risk of developing stroke. Some of the observed motivations gathered from the review include: Stroke responsible for a greater percentage of death worldwide, challenges of quality of service to patients, shortages of medical facilities and complexity of the diseases;

Shortages of medical experts, poor diagnosis, and late detection of the disease. Table 1 shows the summary of the review.

Table 1: Some of the available Predictive Systems for Stroke Predictions

Researcher(s)	Year	Motivation(s)	Method used
Boshra and Mirsaeid [5]	2015	<ul style="list-style-type: none"> - Over 12 million deaths are caused by Heart diseases and stroke in the world annually of which 50% could have been prevented by controlling risk factors. - It was observed that heart disease where stroke is the major cause of it are expected to be the main reason for 35 to 60% of total death expected worldwide by 2025. 	J48 Decision Tree, K-Nearest Neighbour (K-NN), Naïve Bayes (NB), Sequential Minimal Optimization (SMO) of which J48 Decision Tree was adopted due to having the best accuracy result.
Mudasir [4]	2017	<ul style="list-style-type: none"> - Almost 5million persons affiliated with Cardiovascular Disease (CVD) of which stroke is one of the major cause are hospitalized each year. - In 2010, over 16million death caused by CVD and also led to293 million Disability Adjusted Life-Years (DALYs) lost accounting for approximately 30% of all death and 11% of DALYs lost that year. - Between 1990 and 2010; death from CVD increased from 26% to 29% of all deaths globally. 	Multilayer Perceptron Data mining technique.
Oguntimilehin <i>et al.</i> , [16]	2019	<ul style="list-style-type: none"> - Shortage of medical practitioners, medical facilities and complexity of diseases among others gave rise to the increase in mortality rate among patients. 	Partial Tree (PART) technique.
Sabibullah and Kasmir [19]	2010	<ul style="list-style-type: none"> - Stroke is responsible for 3million deaths in developing countries and a major cause in Asian countries. - Estimation of stroke risks in population being not only helpful for healthcare providers but also important to identify persons at elevated risk and to select proper treatments in clinical trials. 	Stacked Topology – Artificial Neural Network (ST-ANN) model
Duen – Yian <i>et al.</i> , [7]	2011	<ul style="list-style-type: none"> - It was seen that cerebrovascular disease was ranked the second or third of the top 10 death causes in Tiawan. - Approximately 13,000 people die every year since 1986 and these cause for a plan on how to bring solution to these menace. 	Decision Tree, Bayesian classifier, and Back-Propagation Neural Network (BPNN).

			Decision Tree was adopted as the best due to its accuracy.
Chaitrali and Sulabha [6]	2012	<ul style="list-style-type: none"> - The challenge of quality of service to patients. - Poor diagnosis which has lead to many disastrous consequence especially death. - Over 17million total global deaths surveyed by WHO are due to heart attacks and stroke. 	<p>Decision Trees, Naïve Bayes and Neural Network.</p> <p>Neural Network was adopted as the best due to high accuracy.</p>
Shadman <i>et al.</i> , [20]	2018	<ul style="list-style-type: none"> - Late detection of cardiac diseases - Poor supervision of clinicians brings about increase in mortality rate. 	<p>Naïve Bayes, Support Vector Machine (SVM), Random Forest (RF), Simple Logistic and ANN.</p> <p>SVM was adopted due to its high accuracy.</p>
Asha and Sophia [2]	2010	<ul style="list-style-type: none"> - To diagnose heart disease was a significant and tedious task to embark upon in medicine 	<p>Naives Bayes, Decision List, and K-NN</p> <p>Naives Bayes was adopted as the best with a high accuracy.</p>
Leila <i>et al.</i> , [11]	2013	<ul style="list-style-type: none"> - Based on studies of more than 56million deaths in 2001, it was found that 7.1 million cases were due to heart disease and 5.4 million were also due to stroke. This indicates that stroke is the second major cause of death in the world that is nearly 10% of all deaths reported. - No studies have been conducted on the C4.5 Decision Tree and KNN regarding stroke rather most of the studies performed on stroke diagnosis and its species differentiation focused on image – processing techniques. 	<p>K-NN and C4.5 Decision Tree.</p> <p>C4.5 Decision Tree was adopted.</p>
Beant and Williamjeet [4]	2015	<ul style="list-style-type: none"> - About 25% deaths in the age group of 25-69 years occur because of heart diseases. In urban areas, 32.8% deaths occur because of heart ailments, while 22.9% in the rural areas. - WHO estimated by 2030, almost 23.6 million people will die due to heart disease. 	Genetic Algorithm
Ashwini and Chandra [3]	2016	<ul style="list-style-type: none"> - The challenge of quality of service to patients. - Failures in some decisions taken by doctors based on their experience in handling some cases. 	<p>Neural Network and Hybrid System.</p> <p>Hybrid System was adopted due to its high accuracy.</p>
Ankita <i>et al.</i> , [1]	2016	<ul style="list-style-type: none"> - Heart disease is the most challenging disease for reducing patient number that 	K-Means and ANN

		<p>is the major cause of death among patients worldwide.</p> <ul style="list-style-type: none"> - WHO has calculated that over 12million death occurs in the every year due to the heart disease. 	ANN was adopted for this work as the best for heart prediction.
Purusothaman and Krishnakumari [17]	2015	<ul style="list-style-type: none"> - Poor Clinical decisions lead to increase in mortality rate. - Incompetent clinicians handling some patients are not good in predicting the heart disease. - Due to long queues and time wastage many lives have been lost. 	<p>Decision Tree, Association Rule, K-NN, ANN, SVM, Naives Bayes and Hybrid Approach.</p> <p>Hybrid Approach was adopted due to its high accuracy.</p>
Ishtake and Sanap [9]	2013	<ul style="list-style-type: none"> - Diagnosing of patients correctly and administering effective treatments have become quite a major challenge. - Poor clinical decisions lead to increase in mortality rate. - High cost of treatment of a patient, not affordable by every patient. 	Decision Tree, Naives Bayes and Neural Network.
Lela <i>et al.</i> , [12]	2015	<ul style="list-style-type: none"> - Out of 15million people that suffer stroke yearly, 6million die and a further 5million are left with permanently disability. - Stroke is believed to be the second leading cause of disability after dementia. 	Artificial Neural Network (ANN)

4. CONCLUSION AND RECOMMENDATION

Looking through the reviewed work, it is clear that stroke has killed millions of people worldwide both old and young ones. The study has also shown that large number of deaths occur annually due to some factors which include late identification of the disease, shortages of medical practitioners and medical facilities, high cost of treatment, time wasting due to many patients on the queue, and poor clinical decisions. The introduction of medical informatics and computer science (Information Technology) into the medical field to help manage some deadly disease like the one discussed in this review is a welcome idea. Based on this new development some computer based predictive models were developed to handle prediction and diagnosing of stroke cases.

Most of the models are based on available symptoms and risk factors. The models are expected to be of great benefits to the masses most importantly patients at risk of developing stroke. During the cause of developing the systems, accuracy and reliability should always be in high consideration to ensure thorough evaluation of it before putting it up to be used by the public. From the reviewed work, it is observed that most of the stroke predictive systems were implemented using various

data mining methods like J48 Decision Tree, K-Nearest Neighbour (K-NN), Naives Bayes, Support Vector Machine (SVM), Artificial Neural Network (ANN), and Hybrid Approach.

It is also observed that most of the adopted methods were based on the number of records, instances and attributes. Use of larger dataset, focus on improved accuracy and making the systems mobile application based that can be easily accessible from any point taking the advantage of fast growing digital age is encouraged for researchers to look into.

ACKNOWLEDGEMENT

Our sincere appreciation goes to numerous authors whose works have been used in this review, Afe Babalola University Multisystem Hospital and Ekiti State University Teaching Hospital, both in Ado-Ekiti where medical experts were consulted. Our constructive criticisms where applicable are meant for nothing but improvement.

REFERENCES

[1] Ankita R.M., Madhuri N.T., and Pooja T.W. **Heart Disease Prediction using ANN and Improved K-Means,**

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering (IJIREICE). vol.4, pp. 221-224, 2016.

[2] Asha R., and Sophia R. **Diagnosis of Heart Disease using Data mining algorithm**, Global Journal of Computer Science and Technology. vol.10, pp. 38-43, 2010.

[3] Ashwini S., and Chandra N. **Different Data Mining Approaches for Predicting Heart Disease**, International Journal of Innovative Research in Science, Engineering and Technology. vol. 5, pp. 277-281, 2016.

[4] Beant K., and Williamjeet S. **Analysis of Heart Attack Prediction System using Genetic Algorithm**, International Journal of Advanced Technology in Engineering and Science. vol.3, pp. 87-94, 2015.

[5] Boshra B., and Mirsaeid H.S. Prediction and Diagnosis of Heart Disease by Data Mining Techniques, **Journal of Multidisciplinary Engineering Science and Technology (JMEST)**, vol. 2, pp. 164-168, 2015.

[6] Chaitrali S.D., and Sulabha S.A. **Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques**, International Journal of Computer Application. vol.47, pp. 44-48, 2012.

[7] Duen-Yian Y., Ching-Hsue C., and Yen-Wen C. **A Predictive Model for Cerebrovascular Disease using Data Mining**. Expert Systems with Applications, vol. 38, pp. 8970-8977, 2011.

[8] Easton J.D. Hauser S., and Martin J. **Cerebrovascular Diseases**, Harrison's Principles of Internal Medicine, New York: McGraw-Hill Publisher, 2001.

[9] Ishtake S.H, and Sanap S.A. Intelligent Heart Disease Prediction System Using Data Mining Techniques. International Journal of Healthcare and Biomedical Research. vol.1, pp. 94-101, 2013.

[10] Jauch E.C., Saver J.L., Adams H.P., Bruno A., Connors A.A., Demaerschalk B.M., Khatri P., McMullan P.W., Qureshi A.I., Rosenfield K., Scott P.A., Summers D.R., Wang D.Z., Wintermark M., and Yonas H. **A Guideline for Healthcare Professionals**, American Heart Association/American Stroke Association. vol.44, pp.870-947, 2013.

[11] Leila A., Reza A., Mohammad T.F., Sayed A. M., F.J, Fariborz K., Rasul N., and Nafiseh T. **Prediction and Control of Stroke by Data Mining**, in *International Journal of Preventive Medicine, 8th Iranian Neurology Congress*, vol. 4, pp. 245 – 249, 2013..

[12] Lela M., Salah A.M., Julian W., and Gillian P. **Artificial Neural Network Model in Stroke Diagnosis**. In *17th UKSim-AMSS International Conference on Modeling and Simulation*, 25-27, 2015, Cambridge, UK.

[13] Moloud A., Sharareh R., Niakan K., Tole S., Much I.S., and Goli A. **Comparing performance of Data Mining Algorithms in Prediction Heart Diseases**. International Journal of Electrical and Computer Engineering (IJECE), vol.5, pp.1569 – 1576, 2015.

[4] Mudasir M.K. **Cardiovascular Disease Prediction Using Data Mining Techniques: A review**, Oriental Journal of

Computer Science and Technology, vol.10, pp.520-528, 2017.

[15] Oguntimilehin A. **A Mobile Malaria Fever Clinical Diagnosis System Based on Non Nested Generalized Exemplar (NNGE)**, International Journal of Emerging trends in Engineering Research, vol. 8, No.2, pp. 259-264, 2020. <https://doi.org/10.30534/ijeter/2020/03822020>

[16] Oguntimilehin A, Babalola Gbemisola. O, Olatunji K.A. **A Clinical Diagnostic Model Based on Supervised Learning**, International Journal of Advanced Trends in Computer Science and Engineering. vol. 8, pp. 949-953, 2019. <https://doi.org/10.30534/ijatcse/2019/94832019>

[17] Purusothaman G., and Krishnakumari P. **A Survey of Data Mining Techniques on Risk Prediction: Heart Disease**, Indian Journal of Science and Technology, vol. 12, pp. 1-5, 2015.

[18] Ramin A., Parham A., and Mohammad R.T. **Heart Disease Diagnosis using Data Mining Techniques**. International Journal of Economics & Management Sciences, vol. 6, pp. 1-5, 2017.

[19] Sabibullah M., and Kasmir R.S.V. **Prediction of Stroke Risk through Stacked Topology of ANN Model**. International Journal of Advanced Research in Computer Science. vol.1, pp. 170-179, 2010.

[20] Shadman N., Rakib R., Rasedul I., and Mohammad H. **Heart Disease Detection by Using Machine Learning Algorithms and a Real – Time Cardiovascular Health Monitoring System**, World Journal of Engineering and technology. vol.6, pp. 854-873, 2018.

[21] Shortliffe E.H., and Blois M.S. **The Computer Meets Medicine and Biology: Emergence of a Discipline**. Medical Informatics - Computer applications in health care and biomedicine. New York: Springer-Verlag, pp.3-40, 2001.

[22] World Health Organization (WHO), **Cardiovascular Diseases**, World Health Organization, 2020. <https://www.who.int/health-topics/cardiovascular-diseases>, accessed 06/07/2021

[23] Bhatiaand N. and Jyoti K. **An Analysis of Heart Disease Prediction using Different Data Mining Techniques**, International Journal of Engineering Research and Technology, Vol. 01, issue 08, 2012.