A Thorough Study on Heart Disease Detection through Machine Learning

Pranav Madihalli
Dept. of Computer Engineering
Modern Education Society's College of Engineering
Pune - 01, India

Sanket Mahajan[2]
Dept. of Computer Engineering
Modern Education Society's College of Engineering
Pune - 01, India

Aishwarya Jadhav[3]
Dept. of Computer Engineering
Modern Education Society's College of Engineering
Pune - 01, India

Saurabh Mirajkar[4]
Dept. of Computer Engineering
Modern Education Society's College of Engineering
Pune - 01, India

Rubeena A Khan[5]
Dept. of Computer Engineering
Modern Education Society's College of Engineering
Pune - 01, India

Mahesh Shinde[6]
Dept. of Computer Engineering
Modern Education Society's College of Engineering
Pune - 01, India

Abstract— Heart health is paramount for the survival of humans as the heart is one of the supreme organs for survival. The incidence of Heart disease and its prevalence has been increasing day by day. This has been attributed to the complicated mechanisms of heart disease and its diagnosis. Heart disease is a complex phenomenon that is caused due to various different attributes. Therefore, the prediction of heart disease has been performed manually by the doctors which takes a lot of time and effort. There has been a collection of researches that have been performed to make heart disease prediction but most of them fall short of their goals in different aspects. The analysis of the past works on this paradigm has allowed for insightful information that is essential for the development of our approach.

Keywords— K-means Clustering, Pearson Correlation, and Deep Belief Network, Decision Tree

I. INTRODUCTION

There has been an increase in the number of Heart diseases in the populous which is also being increased day by day. This increase in the incidences of heart diseases is highly alarming. This is due to the fact that heart diseases are extremely difficult to detect and predict. This leads to an inordinate increase in the heart disease instances that can be attributed to this failure in diagnosing the patient. The patient has to undergo numerous tests and other formalities to complete the diagnostic procedure, which becomes a hassle for the majority of the patient. The doctors are also busy with their schedules and cannot effectively pay attention to a time taking procedure.

The paradigm of heart disease is also aggravated by the fact that most of the sufferers are from the senior citizen category. This can be attributed to the fact that the improvements in the medical field have led to a lower risk of deadly diseases. The enhancements in the medical platform have significantly decreased child mortality rates. This has increased the population considerably. Coupled with the fact that the nutritional improvements and pharmaceutical advancements have lowered disease and risk which has disproportionately decreased the rate of mortality. This has resulted in a large population of elderly individuals.

These are the majority of old persons that usually do not have the strength to go through the diagnostic procedure. There are also patients that are confined to a bed or a wheelchair, which makes it even more difficult to transport them to the hospital every day or for the duration of the diagnostic procedure. These individuals find it difficult to effectively get themselves diagnosed at the hospital which leads to further delay in identifying if there is a presence of any form of heart disease.

Therefore, there is a need for a system that can effectively diagnose and predict the presence of heart disease. Due to the inherent complexity of Heart disease prediction and the complicated approach towards the implementation, an effective approach cannot be easily formulated. There are certain researches that are being performed for this purpose which have utilized various approaches and technological improvements for achieving the goals of heart disease protection. Certain approaches have utilized the paradigm of Data Mining for enabling effective prediction through the utilization of a large amount of data from the patient's medical record.

The Data Mining approaches are limited to the data that is being presented, the effective processing of the data increases the error ratio of the approach which is highly undesirable in such a critical system. Therefore, the data mining approach is not as suitable for the implementation of the heart disease prediction system. The reliability of the system can be significantly increased through the use of machine learning approaches.

The machine learning approaches that have been analyzed in this research have been fruitful in achieving a significant improvement in the prediction of heart diseases. But certain limitations and drawbacks have also been noticed in the past works which can be improved further to achieve even higher prediction accuracy.

This literature survey paper dedicates section 2 for analysis of past work as a literature survey, and finally, section 3 concludes the paper with traces of future enhancement

II. RELATED WORKS

Farrukh Aslam Khan et al. presents a centralized detection for shift mechanism for constant and omnipresent tracking using WBANs. A Markov model-based detection technique is considered, due to the variation in time of the ECG results. The ECG is a data point set usually computed at successive time points spanned at consistent time intervals [1]. The extent of the parts of the electrocardiogram frequency change over time. Hence, for such time-series data, a Markov model is a capital choice. In the proposed system the person's body is connected to various physiological monitoring sensors, and a consolidated PS gathers data for real-time irregular event detection from these sensors. Besides, on the PS, the centralized change detection system operates to secure human peculiar data and to point out irregularities and intrusions ahead of sending it to medical servers for additional determination. Whenever it senses abnormalities or intrusions in the data, the device adds an insistent or exigency tag called an irregular tag to the data. On the receiving hand, it raises an alarm when this tag hits the health management systems, check the footing of the physical data and the related medical staff.

Ahmad Rauf Subhani et al presented a new technique for detecting the level of mental stress by investigating quantitative variations among stress and stress conditions of regulation as well as four stages of conditions of stress [2]. They have educed five characteristics from EEG signals: absolute strength, relative power, consistence, asymmetry of amplitude, and phase lag. The proposed architecture for machine learning has made it possible for the regularity of the characteristics extracted, followed by the collection of features using the operating feature of the receiver (ROC) curve, t-test, and Bhattacharya distance. Three classifiers have been added to the selected features: logistic: Regression (LR), SVM, and Naive Bayes classifiers. To avoid classifiers over-fitting, the validation of the model was provided.

C. M Chethan Malode et al. [3], proposed a framework for heart attack risk identification and categorization. The presented framework is split within four distinct modules they named as soft sets creator, fuzzy rule generator, Support Vector Machine classifier, and decision-maker. To reliably define and classify the danger of heart attack among teenagers, soft set theory and fuzzy law are merged along the SVM classifier. The mathematical technique operated correctly to control the ambiguity in input variables and to decode decision making issues is soft set theory. Fuzzy logic offers results which are noise-proof to problems that are ambiguous, incalculable, and more difficult to compute by generating fuzzy IF-THEN laws by utilizing the logical AND operator.

Aykut Diker et al. presents an intelligent system that helps in identifying ECG signals by utilizing time range, morphological, and discontinuous wavelet transforms characteristics for the purpose to differentiate MI samples from normal[4]. The proposed system framework depends on the genetic algorithm and SVM. The ECG dataset for the mechanism is taken from Physikalisch-Technische Bundesanstalt Diagnostic ECG Database (PTBDB) accessible at Physionet. In totality,23 characteristics derived from anatomy, statistical, and DWT domains were utilized for signal representing.

Mohammed Khwaja et al. [5], present a novel categorization scheme for the recognition of CpG island methylation. The framework extracts the characteristics used for defining methylation from each island and implements collateral characteristics procured from image analogs descriptors. These are combined with a DBN (Deep Belief Network) ML model for the categorization state of the DNA methylation, achieving a precision of 94-97 percent. The method developed is capable of very precise CpG island methylation prediction of a given cell line, area of the genome, and chromosome encountering testing. It builds highly responsiveness, quality, and precision, resulting in an enhanced MCC, are weighed up against the state of the art approaches. Accurate prognosis of DNA methylation has omnipresent inferences plus quick recognition of the disease, via categorizing besides divergence of huge data sets.

Mohammed Saqib et al introduces machine learning approach and deep learning LSTM networks to prognosis the sepsis in EMR records. The crucial symptoms of the patient and the first 24 and 36 hours of the lab test report were used by the authors. The authors utilize the MIMIC3 dataset for implementing machine learning methods. Both models of LR and RF might grasp key characteristics, both bio-markers, and indispensable signs, which are related to sepsis [6]. While the LR model, function weights revealed direct and contrary associations betwixt properties and sepsis forecast, RF exhibited only relative significance.

Rajesh Kumar Tripathy et al. proposed a time-scale analysis based on FBSE-EWT of 12-lead ECGs to recognize MI pathology. The first nine signals of the subband were deemed multiresolution research based on FBSE-EWT associated survey of 12-lead ECG. The skewness and the entropy, the kurtosis, function vectors (FVs) from the 12-lead ECG are evaluated. The Statistical examination of the characteristics discloses that the entropy characteristics are highly critical for MI pathology recognition relative to the role of kurtosis and skewness characteristics of 12-lead ECG [7]. Detection of deep neural networks based on DL-LSSVM was suggested. DNN classifiers with FV entropy work better than VF and Kurtosis asymmetric classifiers. Compared to the industry-leading characteristics of 12-lead ECG greater precision, specificity, and reactivity for MI detection is provided by FBSE-EWT features.

P. Prabhu et al. introduces a deep belief network model for the provision of computational knowledge for the prognosis of diabetes-affected patients with optimum accuracy. The authors used the Pima Indians diabetes dataset for forecasting and analyzing system performance. The normalization method is utilized for dataset preprocessing. Then the prediction model that depends on DBN techniques is designed. The precision of the proposed system is effective than the model depends on other classifiers like decision trees, support vector machine, logistic regression, and random forest [8].

Anjan Nikhil Repaka et al. introduces a framework for heart disease prediction that uses data mining methods to enhance the decision support system. The proposed framework utilizes the naive Bayes classification method for effectively diagnosing heart disease with the use of various attributes and offering appropriate treatments for recovery. The presented techniques include four steps: a collection of a dataset, user registration to the system, Naive Bayes classification, and secure data transfer by implementing AES (Advanced Encryption Standard). The outcome proves the effectiveness of the presented system in identifying heart disease risk factors [9].

Md Rysul Kibria Badhon et al. describes a process for creating a mobile device in which the heartbeat is calculated through an Arduino Uno board, pulse sensor, and AT mega 328p PPG-based microcontroller. The system is enabled to track pulse rate, find out absent heartbeats because of premature ventricular contractions. The whole information is displayed on an LCD. The pulse rate and the absent beat Information is passed to the ESP 8266 Wi-Fi module on a serial basis that transmits the instruction to a website via Message MQTT protocol (Queuing Telemetry Transport). Well timed missing beat detection may be recorded to an individual before a potential congestive heart failure or other flaws in your heart. The heart data can be seen by the physician on the web-page and then treat the patient. [10].

Senthilkumar Mohan et al. introduces a hybrid technique named Forest Random with Linear Model (HRFLM). The main intent of this research is to enhance the Heart disease prediction performance. The HRFLM technique uses all features without any feature selection limitations. The authors perform experiments that are utilized to categorize the characteristics of a machine learning algorithm with a hybrid technique. The authors used a computational strategy in HRFLM with the three rules of association in mining, namely apriori, prediction nature, and Tertius on the UCI Cleveland dataset to discover the variables of heart disease. The available documentation leads to the conclusion that relative to males, females have less risk of heart disease. HRFLM uses ANN with backpropagation, along with 13 clinical features as the input [11].

Amin Ul Haq et al. proposed an distinguishing system using the ML models for classifying HD and healthy subjects. To enhance the classification precision and to lessen the computational time of the prediction system, the authors use subsequent backward selection (SBS) of feature algorithm. Dataset of Cleveland cardiac disease was utilized for assessment of the framework [12]. The 70 percent dataset was utilized for training and the remainder for testing. The evaluation metrics utilized to measure the achievement of the proposed architecture. The test outcome reveals the SBS techniques pick out relevant features and the selected features enhance the accuracy by using the KNN classifier.

III. CONCLUSION

A considerable increase in the number of heart-related ailments and diseases has been noticed in recent years. The heart disease paradigm has also been one of the most fatal diseases that have been increasing in the number of fatalities recently. Therefore, there is an urgent need to provide some improvement in this regard to reduce the number of heart disease-related fatalities through effective prediction approaches. There have been several approaches that have implemented the heart disease prediction paradigm that has been studied with great detail in this research paper. These approaches have been effective in achieving their prescribed goals, but have some drawbacks or limitations that need to be addressed.

REFERENCES

[1] Farrukh Aslam Khan, Nur Al Hasan Haldar, Aftab Ali, Mohsin Iftikhar, Tanveer A. Zia and Albert Y. Zomaya, "A Continuous Change Detection Mechanism to Identify Anomalies in ECG Signals for WBAN-Based Healthcare Environments", IEEE Access (Volume: 5), 09 June 2017.

[2] Ahmad Rauf Subhani, Wajid Mumtaz, Mohamed Naufal Bin Mohamed Saad, Nidal Kamel, and Aamir Saeed Malik, "Machine Learning Framework for the Detection of Mental Stress at Multiple Levels", IEEE Access (Volume: 5), 05 July 2017.

[3] C. M Chethan Malode, K. Bhargavi, B. G Gunasheela, G. Kavana, and R. Sushmitha, "Soft set and Fuzzy Rules Enabled SVM Approach for Heart Attack Risk Classification Among Adolescents", 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 16-18 Aug. 2018.

[4] Aykut Diker, Zafer Cömert, Engin Avci, and Subha Velappan, "Intelligent System Based On Genetic Algorithm And Support Vector Machine for Detection of Myocardial Infarction From ECG Signals", 2018 26th Signal Processing and Communications Applications Conference (SIU), 09 July 2018.

[5] Mohammed Khwaja, Melpomeni Kalofonou, and Chris Toumazou, "A Deep Belief Network System for Prediction of DNA Methylation", 2017 IEEE Biomedical Circuits and Systems Conference (BioCAS), 29 March 2018.

[6] Mohammed Saqib, Ying Sha and May D. Wang, "Early Prediction of Sepsis in EMR Records Using Traditional ML Techniques and Deep Learning LSTM Networks", 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 29 October 2018.

[7] Rajesh Kumar Tripathy, Abhijit Bhattacharyya, and Ram Bilas Pachori, "A Novel Approach for Detection of Myocardial Infarction from ECG Signals of Multiple Electrodes, IEEE Sensors Journal (Volume: 19, Issue: 12, June 15, 15 2019).

[8] P. Prabhu and S. Selvabharathi, "Deep Belief Neural Network Model for Prediction of Diabetes Mellitus", 2019 3rd International Conference on Imaging, Signal

Processing and Communication (ICISPC), 19 December 2019.

[9] Anjan Nikhil Repaka, Sai Deepak Ravikanti and Ramya G Franklin, "Design And Implementing Heart Disease Prediction Using Naive Bayesian", 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 10 October 2019.

[10] Md Rysul Kibria Badhon, Anadi Ranjan Barai, and Fatematuz Zhora, "A Microcontroller Based Missing Heartbeat Detection And Real-Time Heart Rate Monitoring System", 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 04 April 2019.

[11] Senthilkumar Mohan, Chandrasegar Thirumalai, and Gautam Srivastava, "Verifiability Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques", IEEE Access (Volume: 7), 19 June 2019.

[12] Amin Ul Haq, Jianping Li, Muhammad Hammad Memon, Muhammad Hunain Memon, Jalaluddin Khan, and Syeda Munazza Marium, "Heart Disease Prediction System Using Model of Machine Learning and Sequential Backward Selection Algorithm for Features Selection", 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), 29-31 March 2019.