ISSN- 2394-5125 VOL 7, ISSUE 19, 2020

CARDIOVASCULAR DISEASE FORECASTING USING MACHINE LEARNING TECHNIQUES

Shah Samkitkumar Rajnikant [1], Dr. Harsh Lohiya[2]

[1]*Research Scholar, Dept. of Computer Science & Engineering, Sri Satya Sai University of Technology and Medical Sciences, Sehore Bhopal-Indore Road, Madhya Pradesh, India*

[2]*Research Guide, Dept. of Computer Science & Engineering, Sri Satya Sai University of Technology and Medical Sciences, Sehore Bhopal-Indore Road, Madhya Pradesh, India*

ABSTRACT

Humans have an inherent need for healthcare, hence providing it is a mandatory life responsibility. Heart and circulatory system diseases are often referred to as cardiovascular disease. High-risk patients were able to lower their risks thanks to decisions informed by early methods of predicting cardiovascular diseases. Due to the sheer volume of data available in the health care industry, machine learning algorithms are essential for accurately predicting cardiac conditions. The latest studies have focused on finding ways to combine these techniques to create hybrid machine learning algorithms. The proposed research would benefit greatly from the pre-processing of data via measures such as the removal of noise, the removal of missing data, the substitution of default values where appropriate, and the classification of attributes for use in prediction and decision making. The diagnostic model's performance is evaluated using techniques including classification, accuracy, sensitivity, and specificity analysis. This study introduces a methodology for predicting whether a person has heart disease. Accurate models for predicting cardiovascular disease are presented by contrasting the outputs of a number of statistical techniques, including the Support Vector Machine, the Gradient Boosting model, the Random Forest, the Naive Bayes classifier, and the logistic regression technique, on data collected from a given region. To better predict cardiovascular disease, we offer a narrative approach that employs machine learning techniques to identify salient aspects. Different features and multiple classification techniques are proposed to create the prediction model. Through the use of a hybrid random forest and a linear model, we are able to improve the performance of a heart disease prediction model from 80% to 92%.

KEYWORDS: Cardiovascular Disease Prediction, Machine Learning Techniques, Random forest linear model.

INTRODUCTION

The rising cost of healthcare is mostly attributable to lifestyle and genetic factors that have emerged in recent decades. It generates a mountain of information over time. Consequently, the health or survey data are being thrown away. However, this has changed in recent years with the advent of data analytics. Healthcare facilities and non-governmental organisations (NGOs) are using data to their advantage by extracting actionable data. Heart disease now represents the single greatest cause of death in the developed world. The way this illness manifests in the body makes a speedy recovery impossible for those who suffer from it. Timely patient diagnosis is, thus, the most challenging aspect of medical practise. The hospital's poor reputation is due, in large part, to misunderstandings and incorrect diagnoses made by its staff. India raises doubts about the accessibility of the supposedly difficult and financially out-of-reach treatment for this disease. The normal ranges for blood pressure, cholesterol, and pulse rate vary from person to person. Medical studies show that a blood pressure of 120 over 80, cholesterol of 200, and pulse rate of 72 are all considered normal. According to the WHO, cardiovascular world is the leading cause of death worldwide, accounting for more than 12 million deaths annually. It's a devastating disease that's causing even more problems in India. The technique through which the unhealthy condition is investigated is intricate. A precise and accurate measurement is required. Because patients in some areas are put in jeopardy due to a lack of specialists. Typically, cardiologists are the ones to make this diagnosis (Who generally treat the heart disease patients). Integrating these methods into the healthcare information infrastructure has tremendous potential for improvement. In order to determine which machine learning approach is most suited for integrating the survey data collected in June, we must first compare and contrast many techniques. In order to predict the uncertainty levels of cardiovascular diseases based on the qualities available, this research presents the various machine learning techniques. The medical data sets included come from studies that piqued interest in different parts of the world. Machine learning is the practice of getting the most out of a system without needing to explicitly programme it. They are put to use in examining the analytical structure of high-dimensional, multifaceted data sets, such as those pertaining to cardiovascular diseases. They are employed for the recognition of patterns that aid in the study and administration of forecasts and controls.

According to the World Health Organization, the leading causes of death in the world are cardiovascular disease and chronic respiratory disorders. There are over 17.5 million deaths each year caused by cardiovascular illnesses (CVDs), and another 1.59 million caused by chronic respiratory disorders. Chronic respiratory disorders include asthma, occupational lung diseases, hypoxemia, and hypercapnia; chronic cardiovascular diseases include heart attack, stroke, heart failure, arrhythmias, and valve issues. The insufficiency of resources across healthcare systems slows the rate at which diseases can be treated. Therefore, people have lost faith in the healthcare system as a whole. The capacity to quickly respond to a patient's critical state is beneficial for both the patient and the clinician. Researchers from all across the world are trying to find better ways for medical healthcare systems to reduce unnecessary deaths. Modern medical methods help healthcare systems by making prognostic inferences about a patient's health status using a variety of machine learning methods (e.g., Naive Bayes, random forest, and neural networks). These methods look at a patient's history in order to foresee potential health issues. Our current understanding is that these methods can only make inferences regarding a patient's state of health. They are not, however, designed to foretell a patient's prognosis in any way. In emergency situations, time is of the essence because a patient's health might worsen quickly. Typically, first responders arrive at the scene of an incident within three minutes. During that time, doctors may have a hard time treating patients since they don't know enough about their conditions. When caring for a patient, we believe that the ability to foresee their future needs is essential. The ability to foresee a patient's future condition could save lives and give medical staff valuable insight. By developing a system that can anticipate a patient's state in the next 60 seconds to 3 minutes based on a patient's history of cardiovascular and chronic respiratory disorders, this prediction helps alleviate this issue. This device can detect early warning signals of chronic diseases like asthma and heart disease and send out critical health alarms. Then, each health alert is placed into one of 38 categories, and an estimated patient health score is calculated. The dataset used contains vital sign readings from a variety of patients following 32 different types of surgery. We have ranked the 10 most significant vital signs for each disease with the input of doctors. These measurements are then fed into a linear regression and polynomial regression model with degrees 2, 3, and 4 to produce a 60-second and 3-minute vital signs forecast. There are between 50,000 and 100,000 samples of vital signs utilised to train these regression models, all taken within a 10-ms window. Classifiers such as the Support Vector Machine, the Decision Tree, and the Naive Bayes classifier are fed with these anticipated values to make predictions about the patient's future health. Results showed the decision tree worked very well as a forecasting tool. This evidence supports the conclusion that the proposed approach is highly effective in predicting the long-term outcomes of patients suffering from cardiovascular and chronic respiratory disorders.

LITERATURE REVIEW

Divya Annepu et al (2019), A Heart disease has surpassed all other causes of death in recent years, for both men and women. Due to heart attacks, the death rate in India has risen to 24.8%. Predicting the likelihood of developing cardiovascular disease is a crucial first step in reducing the severity of the problem. The prediction of heart diseases can be automated, which will save time and effort and help reach this goal. Machine learning, a relatively new field that makes use of data mining, has been shown to be particularly useful for predicting cardiovascular diseases. All relevant patient medical patients is gathered, organized, and used to train the dataset. To better comprehend medical data and avoid cardiovascular diseases, mining is utilized to detect patterns that were previously concealed and not discovered due to unknown links. The accuracy of our python-based heart disease prediction is 97.56% thanks to the usage of Random Forest classifiers.

Munaga Meghana et al (2020), Cardiovascular disease ranks high on the list of global health concerns. The clinical data analysis field has a significant challenge in the area of cardiovascular disease prediction. Hybrid machine learning (ML) has shown to be a useful tool in helping the healthcare industry and hospitals make sense of the vast amounts of data they generate. Furthermore, recent innovations in various IoT domains have included the application of ML techniques (IoT). Predicting heart disease using ML techniques is only partially illuminated by existing research. To better predict cardiovascular disease, we offer a narrative approach that employs machine learning techniques to identify salient aspects. Different features and multiple

ISSN- 2394-5125 VOL 7, ISSUE 19, 2020

classification techniques are proposed to create the prediction model. Through the use of a hybrid random forest and a linear model, we are able to improve the performance of a heart disease prediction model from 80% to 92%.

RESEARCH METHODOLOGY

Understanding a patient's susceptibility to heart disease may help doctors decide on the best course of treatment. The extinction of the human race could be aided by this. In this work, a prognostic and classification model is used to assess the health of patients with cardiovascular and chronic respiratory (hypoxemia and hypercapnia) illnesses. We analysed each disease's vitals using data collected from actual patients. First, linear and polynomial regression models are used to vital signs to make predictions for the next one and three minutes. By applying the categorization models to the expected results, we have been able to place each vital sign into one of three categories: low, normal, or high. The researchers discovered a hybrid model that, given a minute's worth of data, could predict vital signs and classify into a distinct label, so facilitating the healthcare provider's ability to make decisions more swiftly. Data from many time series can be used in conjunction with the forecasting model to improve accuracy and performance significantly above previous efforts. In order to accurately anticipate the patient's future condition, the model is sensitive enough to isolate optimal hybrid regression models. Health of the Patient The suggested system falls under the larger category of telehealth, specifically PMC telehealth. The major purpose of the PMC TeleHealth system is to present real-time medical data about a patient, such as vital signs, graphical charts, health alarms, etc. This information will help doctors and other medical professionals provide more individualized care. The PMC Telehealth system's most notable feature is its ability to foresee medical emergencies by analyzing collected vital signs data. The CCVS's remit includes predicting the onset of diseases. This study intends to boost the CCVS's ability to detect early warning indications of critical illnesses like hypoxemia, hypercapnia, and heart failure, which is necessary if healthcare infrastructures are to keep up with escalating demand.

Support Vector Machine

It is a supervised machine learning technique that uses labeled target tuples as training data to determine the best hyper-plane. The objective is to locate the hyper-plane of class intersections. As there may be more than one hyper-plane that can find this, finding the one with the largest margin that optimizes distance between classes is the objective. Using the aforementioned hyper-plane, the newest piece of data to be classified can be classified with ease.

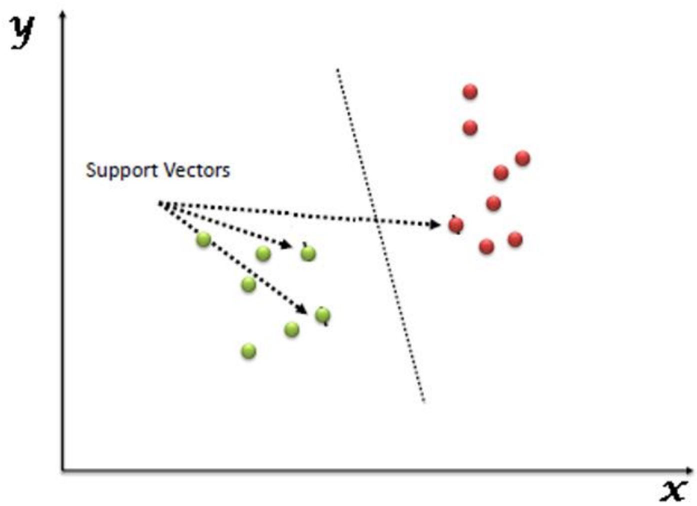

Figure 1.1 Support Vector Machine

K – nearest neighbor

The KNN model, developed by Fix and Hodge, is one of the simplest and most widely implemented supervised classification algorithms available today. To find its nearest neighbours, it works the distances between test and training data points. Once a new sample is collected, it is placed in the category that most closely matches its nearest neighbour. When using KNN, K stands for the number of neighbors. This 'K' value plays a crucial role in the classification. Numeric data, including those measured in Euclidean, Manhattan, and Minkowsky coordinate systems, are suitable for KNN's use.

ISSN- 2394-5125 VOL 7, ISSUE 19, 2020

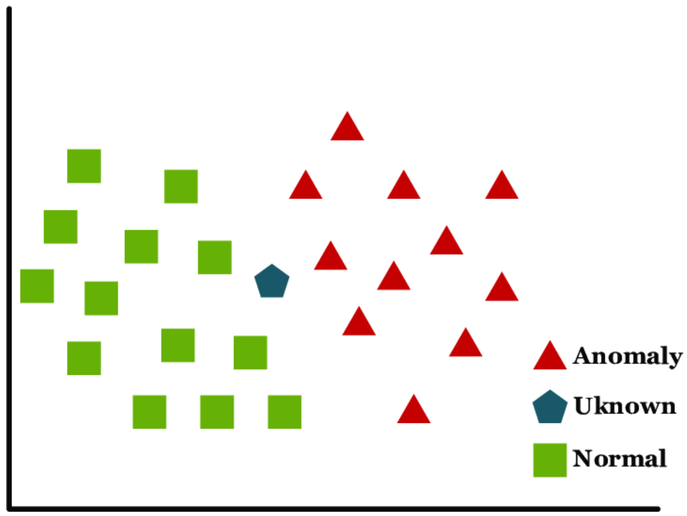

Figure 1.2 K- Nearest Neighbour

Decision Tree

When it comes to supervised learning techniques, Decision Tree is one of the few that can deal with both continuous and categorical data. It kicks off the classification procedure by splitting the dataset into two or more groups with shared characteristics based on the most important predictors. Each characteristic's entropy is then determined. The property with the highest information gain or lowest entropy is chosen as the dividing factor. Recursion applies the same method to the remaining attributes.

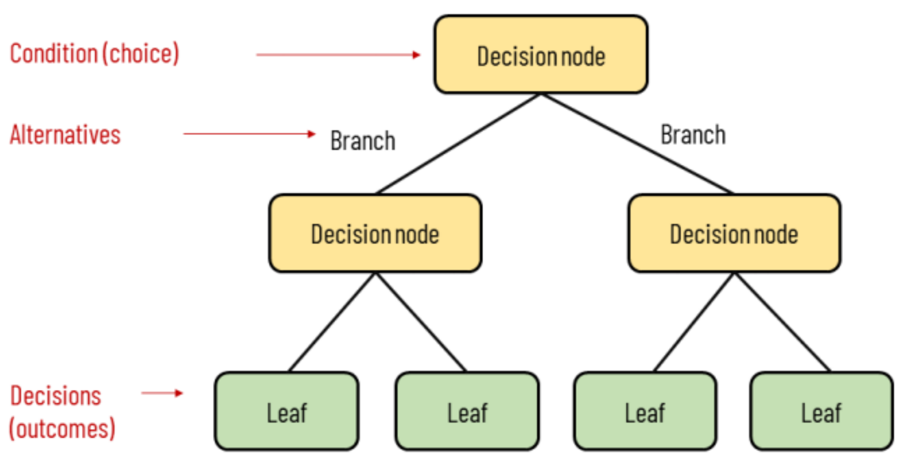

Figure 1.3 Decision Tree

ISSN- 2394-5125 VOL 7, ISSUE 19, 2020

Artificial Neural Network

An ANN, or "neural Network," is a system that is designed to mimic the functioning of the human brain (NN). The system mimics the neurons in the human brain in order to execute calculations and numerical models on multiple levels. It performs the same function as a single neuron in a human brain. It is multi-tiered, with different perceptrons at each stage. Putting a value on each perspective affects the final result. It also features a number of input layers responsible for computation and weight redistribution, as well as an output layer responsible for generating the actual output.

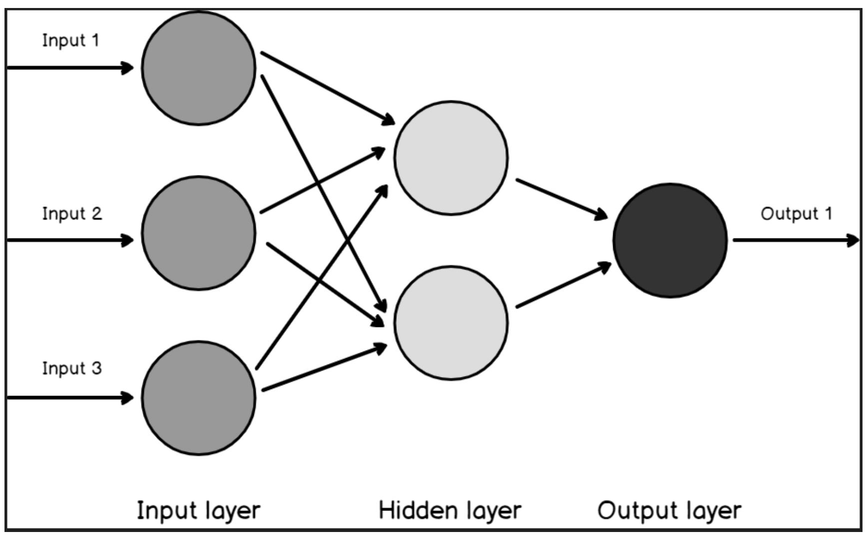

Figure 1.4 ANN with Input Layer

Data Source

There is a large amount of patient information stored in healthcare databases. The term "heart disease" is commonly used to refer to a wide variety of illnesses that are detrimental to human cardiac heart. Cardiovascular disease ranks among the top diseases in the animal kingdom. Cardiovascular disease refers to a group of disorders that affect the cardiovascular system, including the heart and arteries, and hence the body's ability to pump blood and distribute it throughout the body. Information on heart disease was obtained from records in Cleveland, Hungary, Switzerland, and the Long Beach VA (UCI machine Learning Repository). Separating disease-related patterns into individual datasets. Two distinct collections of records, a "training dataset" and a "testing dataset," are generated. There were a total of 920 records obtained, each with 76 medically-related features.

Analysis of Data

Given that the dataset has blanks and duplicates, this stage is primarily responsible for executing data preprocessing operations such as cleaning, integrating, filling, and deleting. Fault prediction is the result.

Operating Environment

R is a programming language built for statistical computing and visual representation; it is therefore well suited for use in data analysis. An assortment of packages for statistical computing and graphical representation makes it easy for the user to do analyses and see the findings visually, paving the way for the creation of a prediction system tailored to the user's needs. R is free, cross-platform programme that may be run on either UNIX or Windows. If you're looking for accurate prediction outcomes, R is your best bet. Heart disease can manifest in a variety of ways, but everyone has a unique set of heart variables that determine whether or not they will develop the disease. The presence or absence of heart disease can be predicted by looking for certain common indicators.

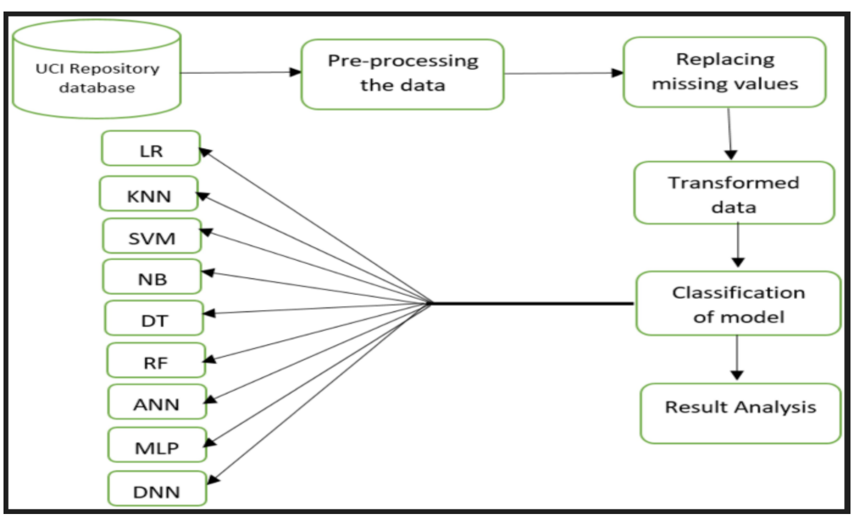

Figure 1.5 Proposed Systems

ISSN- 2394-5125 VOL 7, ISSUE 19, 2020

Decision trees are an example of supervised machine learning algorithms. It can process both numerical and category information. It takes in a set of criteria and returns a binary result (Yes/No, True/False, 1 or 0). Decision tree Classification is a popular approach for managing healthcare data sets. There are some models between this model's output and that of others, such as the knn model and the SVM model. The conditional dependencies of the independent variables determine the orientation of the output lines, which might be horizontal or vertical. This algorithm has a significantly better degree of accuracy than its competitors. The model's examination of the dataset in a tree structure is the key to the algorithm's superior accuracy. This means that the entire dataset has been scrutinized. That's why this model has a better rate of accuracy. The data is analyzed using a tree structure in this model. The plan of action is laid down in a tree format.

RESULTS AND DISCUSSION

Regression techniques: Let us review the findings and compare them to the various prediction models that were provided. A more accurate description of the potential of the regression method can be obtained by thinking about the comparable value of mean absolute percentage error. The performance of a forecasting model is measured by its mean absolute percentage error (MAPE). In this section, we provide MAPE ratings for each regression model alongside their relevant KPIs. Forecasts of cardiovascular disease vital signs over the next minute using linear regression and polynomial regression models are both quite accurate.

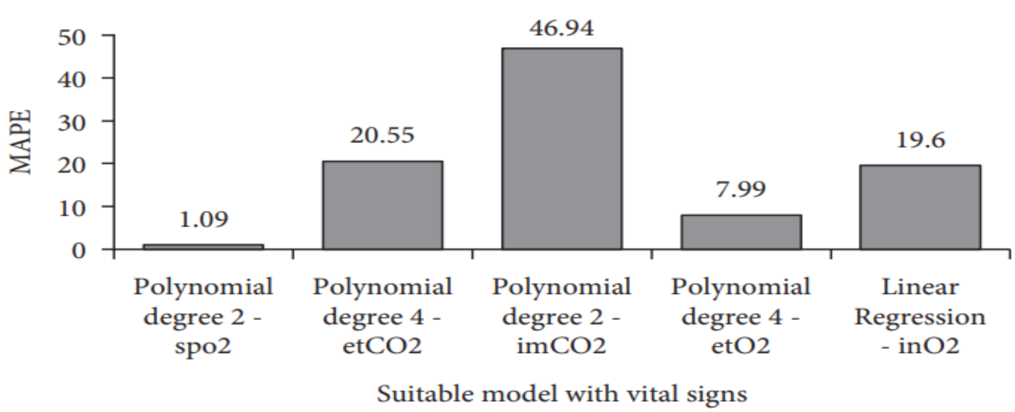

Figure 1.6 Best regression models for 60-second forecasted vital signs.

ISSN- 2394-5125 VOL 7, ISSUE 19, 2020

Classification techniques: We've decided on using Naive Bayes, support vector machine, and decision tree classifiers to divide patients' illnesses into 38 categories. All of the classification techniques have been tested and evaluated by industry standards (i.e., precision, recall, and F-measure). Classification results as a whole have been evaluated using the F-measure, which is the harmonic mean of precision and recall. The results of these evaluations are shown in Table 21. In Table 21, the highest possible values for evaluation measures are shown in yellow boxes, while the lowest possible ratings are shown in pink. The results of using different harnessed classifiers are shown in Table 21. Naive Bayes, Support Vector Machine, and Decision Tree all produce similar results when asked to predict the next 60 seconds for a patient with heart failure: 0.72, 0.82, and 0.75 for precision, 0.67, and 0.60, and 0.61 and 0.80, 0.76, and 0.73 for recall and F-measures. This suggests that we can predict the heart failure patient's status for the following 60 seconds using either Naive Bayes or a decision tree. Similarly, the precision, recall, and F-measure for the next 60 seconds for patients with Chronic Respiratory Disease are 0.30, 0.29, 0.28, and 0.31, 0.31, 0.30. The outcomes of this disease strongly support the use of the Decision Tree in the prediction of chronic respiratory diseases.

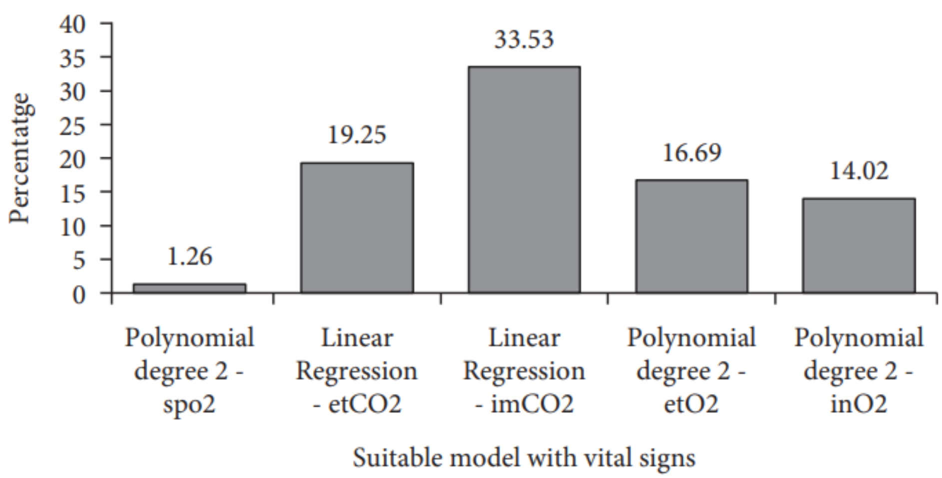

Figure 1.7 Best regression models for 3-minute forecasted vital signs

When asked to forecast the survival prospects of heart failure patients in the next three minutes, Naive Bayes, Support Vector Machine, and Decision Tree all yield very comparable results: 0.59, 0.67, and 0.60 for accuracy, 0.73 for recall, and 0.75 for F-measures. Even if cardiac failure

ISSN- 2394-5125 VOL 7, ISSUE 19, 2020

is suspected, a decision tree can properly forecast the patient's status over the next three minutes. The precision, recall, and F-measure for the subsequent 3 minutes are 0.25, 0.26, and 0.24, respectively, for Chronic Respiratory patients. Predictions of chronic respiratory disorders should be made using either the Naive Bayes or the Decision Tree model, according to these results (for the next 3 minutes values of vital signs). The decision tree's performance has been consistent across all of the tests. The best classification results have been obtained for predicting the status of a cardiovascular patient in the next minute. The Naive Bayes and Decision Tree models have achieved the highest values of precision, recall, and F-measure when applied to the prediction of cardiovascular disease (i.e., heart failure) during the next 60 seconds and 3 minutes. Classifying chronic respiratory disorders, however, has not shown many promising results. Our findings indicate that the proposed system is useful for forecasting the health of cardiovascular patients by highlighting when they are likely to need immediate medical attention (i.e., 60 seconds and 3 minutes).

CONCLUSION

There are many diseases in health care, and cardiovascular disease is one of them. Both diseases have a global impact on mortality because of the modern non-stop lifestyle. A heart attack can occur with no warning. Vital sign predictions made using this intelligent technology may save the life of a patient suffering from cardiovascular and chronic respiratory disease. It is possible that the computer-aided system will help reduce the number of fatal heart attacks and consequent deaths. It is not an easy process to determine which machine-learning model best fits a given set of data. In this research, we've outlined the diseases of a machine-learning-based system for chronic disease prediction, discussing models for forecasting vital sign values. The study's secondary objective is to improve the quality of care provided to in-house patients, particularly those who have recently undergone surgery, by assisting nurses and doctors in getting to them as soon as possible with life-saving medical attention. The prediction model's viability must be evaluated in light of the specifics of the data. In prediction, this work's main findings include demonstrations of the usefulness of polynomial regression in predicting the vital signs used, as well as evidence of the curvilinear character of the vital signs employed. We've done what's needed to use a sizable, all-encompassing dataset to give us faith in our polynomial regression

ISSN- 2394-5125 VOL 7, ISSUE 19, 2020

prediction model. Additionally, the classifiers are trained to determine the patient's condition based on the prediction outcomes. According to our findings, the Decision Tree is capable of accurately categorizing the patient's condition. The decision-tree based classifiers are widely regarded as natural to use. They don't need a plethora of variables, but their application in the context of missing data needs to be scrutinized.

REFERENCES

[1]. Rairikar, A., Kulkarni, V., Sabale, V., Kale, H., & Lamgunde, A. (2017, June). Heart disease prediction using data mining techniques. In 2017 International Conference on Intelligent Computing and Control (I2C2) (pp. 1-8). IEEE.

[2]. Gandhi, Monika, and Shailendra Narayan Singh. "Predictions in heart disease using techniques of data mining." In 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), pp. 520-525.IEEE, 2015.

[3]. Chen, A. H., Huang, S. Y., Hong, P. S., Cheng, C. H., & Lin, E. J. (2011, September). HDPS: Heart disease prediction system. In 2011 Computing in Cardiology (pp. 557-560). IEEE.

[4]. Aldallal, A., & Al-Moosa, A. A. A. (2018, September). Using Data Mining Techniques to Predict Diabetes and Heart Diseases. In 2018 4th International Conference on Frontiers of Signal Processing (ICFSP) (pp. 150-154). IEEE.

[5]. Sultana, Marjia, Afrin Haider, and Mohammad Shorif Uddin. "Analysis of data mining techniques for heart disease prediction." In 2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), pp. 1-5. IEEE, 2016.

[6]. Al Essa, Ali Radhi, and Christian Bach. "Data Mining and Warehousing." American Society for Engineering Education (ASEE Zone 1) Journal (2014).

[7]. Shetty, Deeraj, Kishor Rit, Sohail Shaikh, and Nikita Patil. "Diabetes disease prediction using data mining." In 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), pp. 1-5. IEEE, 2017.

[8]. Methaila, Aditya, Prince Kansal, Himanshu Arya, and Pankaj Kumar. "Early heart disease prediction using data mining techniques." Computer Science & Information Technology Journal (2014): 53-59.

[9]. Dewan, Ankita, and Meghna Sharma. "Prediction of heart disease using a hybrid technique in data mining classification." In 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 704-706. IEEE, 2015.

[10]. M. Fatima and M. Pasha, "Survey of machine learning algorithms for disease diagnostic," Journal of Intelligent Learning Systems and Applications, no. 1, pp. 1–16, 2017.

[11]. H. H. Rashidi, N. K. Tran, E. V. Betts, L. P. Howell, and R. Green, "Arti- ficial intelligence and machine learning in pathology: the present landscape of supervised methods," Academic pathology, vol. 6, pp. 23–74, 2019

[12]. S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, "Can machine-learning improve cardiovascular risk prediction using routine clinical data?" PloS one, vol. 12, no. 4, Article ID e0174944, 2017.

[13]. S. Goldberg, H. M. Ollila, L. Lin et al., "Analysis of hypoxic and hypercapnic ventilatory response in healthy volunteers," PLoS One, vol. 12, no. 1, Article ID e0168930, 2017.

[14]. K. Ara Shakil, S. Anis, and M. Alam, Dengue Disease Prediction Using Weka Data Mining Tool, arXiv e-prints, 2015.

[15]. F. Jabeen, M. Maqsood, M. A. Ghazanfar et al., "An iot based efficient hybrid recommender system for cardiovascular disease," Peer-to-Peer Networking and Applications, vol. 12, no. 5, pp. 1263–1276, 2019.