An Analytical Study On Prediction Of Heart Failure Through Machine Learning

Shreya Yadav, M.S. Takalikar

Abstract—Due to unhealthy lifestyle practices and enormous stress in today's world, it is being completely unpredictable when a person can get a heart attack or heart failures. Even most of the times doctors and health experts are also not able to predict the heart failures, because of this prediction of heart failure remains as a mystery even though on having so much advance technologies in the medical field. Machine learning algorithms also jump into this to predict the heart failures using some techniques like complex event processing and others. These processes involve a large amount of data to learn about the conditions and then they predict the heart failures. It is not possible to provide large amount of data every time to predict the heart failures. So some techniques should be there to provide the same in a moderate amount of datasets condition. So we try to concentrate mainly on evaluating the existing methodologies and try to find the flaws in them so that a new effective idea for heart failure prediction can be brought into light that works efficiently.

Keywords—Congestive Heart Failure Prediction, Machine Learning, Heart Disease.

1. INTRODUCTION
The heart is one of the most essential organs in the human body. It is one of the rare muscles in the body that is constantly working, since before a human is born and usually stops when the human is dead. It is one of the strongest muscles and is admired for its endurance. The Heart is very significant for humans and has also been characterized as the center of all emotion and affection portrayed by the human being. It is very essential and an irreplaceable part of human life, even more so that the brain, which is actually the most important part and also the center of all emotion in a human being. The Human heart is about the average size of that particular human’s fist, very proportional and has some critical responsibilities to fulfill. The most crucial responsibility is fulfilled by the heart is that of pumping freshly oxygenated blood throughout the body and replaces the bad blood or deoxygenated blood from the various parts of the body. This is achieved with the help of the pumping action of the blood. The heart itself is segregated into 4 parts where different blood is stored and pumped into and out of the blood vessels. The bad blood is taken to the lungs to remove the carbon dioxide and fix the oxygen in the blood through a pigment found in the red blood cells called hemoglobin. This pigment is responsible for carrying oxygen in the blood and transporting the carbon dioxide and a by-product of respiration. Oxygen is highly useful for all the different calls in the body and the heart is responsible for providing it to the various parts of the body. Due to lifestyle changes and various other factors such as diet and environment can have a negative impact on the heart in the long run. Due to a sharp increase in the lifespan and life expectancy of humans, the hearts are actually working for a longer time and with a changed lifestyle, they are subjected to massive stress and strain in the lifetime of a human being. This is highly evident in the aging demography now as most of them are suffering from various forms of heart diseases. The most common ailment that is being noticed in the elderly nowadays is the Congestive Heart Failure, it is one of the most debilitating heart diseases that can significantly reduce and hamper the blood flow of the heart to the various parts of the body. The Congestive Heart Failure leads to a weak heart that is unable to pump blood effectively leading to the blood not being able to keep up with the metabolism of the body. The Congestive Heart Failure is one of the most commonly found Heart Ailments in the elderly populous and it is highly dangerous for them. Most of the fatal heart failures are a result of Congestive Heart Failure that results in the loss of blood to the extremities of the body and causes blood perfusion of the organs and the tissues. This leads to very painful congestion of the circulation and pulmonary system which leads to bad blood being reused and further damaging the body. Therefore, it is essential to identify and classify Congestive Heart Failure. It is essential for the survival of the human body as the heart-stopping its function is considered as death. It is also one of the most active muscles in the body that never takes rest. But various lifestyle choices and diseases can lead to the heart getting weakened. Also, the tiny capillaries that provide the heart walls and muscles with fresh oxygenated blood also susceptible to clogging due to cholesterol and other debris being accumulated over a period of time. This debris can lead to reduced supply of blood to the heart which weakens the heart a lot more. All of these conditions add up and ultimately be fatal. The problem is that most of these conditions are highly subtle and can be easily missed even by professionals. This where Machine Learning is instrumental, as it can help isolate the symptoms and predict if that particular person will suffer a heart failure with great accuracy. A lot of research has been done in this field of predicting heart failure with varying amount of success. Most prominently, the researchers utilized the ECG or Electrocardiogram to determine if there are any inconsistencies that can be identified through the complete analysis. The technique is to utilize the various different frequencies in the ECG to identify any abnormalities that are observed. One of the most common technique id to utilize wavelet transformations to identify any morphological changes in the waves of the ECG that indicate the chances of a Congestive Heart Failure. Researchers have also utilized a combination of cardiopulmonary exercises for the purpose of prediction. The authors processed a linear regression of the peak heart rate to estimate the conditions of congestive heart failure. The parameters such as peak heart rate and peak oxygen consumption can be utilized and processed through linear regression to yield a prediction about the patient quite accurately. The most recent example of the prediction models that have been researched extensively in the Heart Variability measures extracted from a patient. The various different parameters such as time-domain measures and the Resting rate intervals, which is the interval between the two R waves formed in the resting phase of the ECG. These intervals and other parameters are then extensively analysed through Support Vector Machine and the system predicts if
the patient is at the risk of a Congestive Heart Failure. Some researchers also focus on reducing the readmission of a patient suffering from Congestive Heart Failure. This is due to very high readmission rates that are noticed in such patients. Therefore, predicting if the condition of the patient would turn worse after discharge would be highly useful in the prevention of a lot of damage that can be done due to this disease. The researchers take into consideration various factors such as social factors as well as medical severity to power the discriminant analysis to form the baseline. Then association rule mining is used to deduce the probability that the patient will be readmitted. Prediction plays a vital role in prevention of the Congestive Heart Failure with the help of Machine Learning. In this paper, section 2 is dedicated for literature review of past work and Finally Section 3 concludes this paper.

2. LITERATURE REVIEW

B. Hu [1] states that for the diagnosis of heart disease the electrocardiogram heartbeats are one of the most essential components. The researchers in this paper utilized the HRV or Heart Rate Variability to accurately detect and classify the Congestive Heart Failure conditions in a patient with the help of an electrocardiogram. They focused on detection using different time scales of HRV. Usually existing system were using only one or two types of time scale. But in this paper they used seven time scales. SVM classifier was used for the classification purpose. The authors have tested this technique extensively and achieved an average of 96% accuracy. The major drawback of this technique is the low accuracy achieved while classification.

W. Pan presents an innovative technique for the purpose of diagnosis of a dangerous heart condition, known as Congestive Heart Failure. The researches [2] noticed that the Autonomic Nervous System (ANS) also shows some signs which can be used to detect the condition of a Congestive Heart Failure. The Multi-Frequency Component Analysis (MFCA) method was proposed to investigate the rhythms changes and complexity of ANS. For that method the input is multi-frequency components such as entropy and utilized them as indices for the Heart Rate Variability for the assessment of Congestive Heart Failure. The technique achieved a classification accuracy of 86%. The main drawback of this technique is the very low accuracy achieved by this technique.

E. Mbazumutima expresses that there must be certain technique that can be utilized for the assessment of patients with Congestive Heart Failure. The authors in this paper have been utilizing the data obtained from the early stages of cardiopulmonary exercise testing. The researchers in this paper have utilized the various parameters in the cardiopulmonary exercise testing such as, peak oxygen consumption rate and peak heart rate to analyze and diagnose Congestive Heart Failure in a patient. The major drawback in this technique is that it has not been tested extensively to assess its impact. [3]

H. Wendt explains that there has been a lot of indices or parameters that have been utilized for the process of classification and identification of the temporal dynamics of the HRV or Heart Rate Variability for the purpose of identification and diagnosis of the Congestive Heart Failure. Due to the fact that Congestive Heart Failure is one of the deadliest heart ailments and timely diagnosis is one of the key reasons for the survival of the patient [4]. Therefore, for this effect, the authors have presented a novel framework that utilizes Wavelet p-leader non-gaussian Multiscale expansions for the purpose of Congestive Heart Failure Identification.

A. Windmon elaborates on the concerns surrounding the COPD or Chronic Obstructive Pulmonary Disease and CHF or Congestive Heart Failure, as they are one of the deadliest heart ailments and are a leading cause of death in the elderly populous of late. Both of these conditions have very similar symptoms which can be utilized to prevent or detect the signs of a CHF or COPD in the early stages [5]. Therefore, the authors have proposed TusisWatch which is a smartphone system that can be used to identify the early symptoms of CHF and COPD. The major drawback in this technique has been due to the large computational complexity and the occurrences of false positives.

D. Destiani states that Congestive Heart Failure is one of the most dangerous heart ailments that can be suffered by a human being. Congestive Heart Failure is also one of the leading causes of heart abnormalities as the heart muscles are strained and eventually deform under the stress. This weakens the heart and it cannot pump blood efficiently across the body, this can be detected by an Electrocardiogram [6]. The authors in this paper propose an innovative technique for the analysis of an ECG by studying the line search and wavelet with the help of a modified Polak-Ribiere algorithm. The proposed technique has a quite low accuracy at 87%.

A. Khayyat expresses that there has been an increase in the number of Congestive Heart Failure patients recently. The Congestive Heart Failure is one of the most debilitating ailments of the heart, characterized by the fact that it causes less blood flow to various organs of the body, which causes swelling and congestion in the body tissues and lungs. Therefore, to help manage this disease efficiently, the authors have implemented a Decision Support Tool for reducing the possibility of readmission into the hospital and also estimate the condition of the patient accurately [7]. The decision support tool has not been experimented extensively to achieve a performance analysis.

S. Potturu explains that the respiratory system is one of the most affected systems for a person suffering from a Congestive Heart Failure. As the respiratory system is one of the most affected organs of the body for a Congestive Heart Failure as it is one of the most important organs of the human body [8]. The respiratory system is responsible for the oxygenation of the blood, if this system is compromised, there will be a group of oxygen levels in the blood. Therefore, the authors have implemented a technique that analyses the respiratory system for signs of CHF and can help identify the impact correctly. The major drawback is that this system has only been conceptualized as of now.

G. Valenza presents an innovative technique for the Spectral Analysis of HRV or Heart Rate Variability, to effectively analyze and diagnose Congestive Heart Failure. This is due to the fact that the Congestive Heart Failure has a lot of impact on the sympathetic and the parasympathetic nervous system [9]. This impact is quite evident in the EEG or Electroencephalography tests and can be quantified effectively. Therefore, the authors in this paper propose a highly accurate technique based on Electroencephalography that is used to automatically classify and identify the patients with Congestive Heart Failure. The major drawback in this
paper is that this is a study and the technique has not been deployed yet.

A. Nainwal [10] states that there are a lot of conditions that cause heart failures in humans. There has been a significant increase in the number of cases involving heart failure in humans. Most of these incidents have been reported to occur in elderly individuals, as an aging population is one of the most vulnerable to Heart Failure. The authors in this study implement a technique that extracts morphological features from an Electrocardiogram to diagnose Congestive Heart Failure in patients. The major drawback of this technique is the high time complexity required to process the Electrocardiogram.

Mingjie Qian proposed a new Temporal Reflected Logistic Regression (TRLR) model that fulfills the two objectives. The TRLR model can utilize as many variables or features as possible from existing EHR data and is more adaptive in that model parameters can be updated by an effective and efficient algorithm for parameter estimation as more data or observations become available[11]. The authors formalized the parameter estimation problem by utilizing the MLE (Maximum Likelihood Estimation) principle. The major drawback of the proposed system is that this is based on two assumptions: first that a patient with HF will eventually die and second is that the survival probability of an HF patient is monotonically decreasing concerning time. Afef Mdhaffar presents a new health assessment approach for heart failure prediction. The proposed approach is called CEP4HFP (Complex Event Processing for Heart Failure Prediction) [12]. It depends on the use of the methodology of Complex Event Processing (CEP), merged with statistical approaches. CEP4HFP (1) collects health parameters, (2) processes collected data by executing analysis rules and (3) triggers alarms if a heart failure has been detected. The used analysis rules are typically based on comparing extracted health parameters to thresholds. In the proposed technique thresholds are automatically calculated and updated at runtime, via the use of threshold customization approach. The major drawback of the proposed technique is that this is a statistical method and it depends on the historical data to calculate a threshold value.

Chi-Jim Chen [13] proposed a system module that can automatically classify HF patients into four different major categories and predict the survival rates for each patient. By predicting and analyzing deterioration patterns of HF diseases, the proposed system can provide prognosis advice and reliable healthcare treatments. The authors applied the SVM tooiset library support vector machines to predict patient’s survival periods based on their features in medical records. The major drawback of the proposed system is that the SVM only really covers the determination of the parameters for a given value of the regularisation and kernel parameters and choice of kernel and it required more training time for large datasets.

Muhammad Saqlain proposed a classification technique for the prognosis of HF diagnosis patient’s survival risk within the time span of one year. The authors collected the data from the AFIC (Armed Forces Institute of Cardiology), Pakistan. They collected the data for only those patients whose treatment was completed within 1 year. Later they applied Naïve Bayes (NB) Classification algorithm on the preprocessed dataset using Rapid Miner [14]. The major drawback of this paper is that the authors missed lots of buried important information from the patient unstructured reports.

Evanthia E. Tripoliti [15] proposed an automated technique for the prediction of adverse events related to HF, utilizing information from saliva and breath biomarkers. The whole technique is distinct into three stages. First stage is preprocessing of data in which missing values are addressed, second is feature selection in which the identification of features that can act as discriminators between the two expected situations (presence of an event or not) are selected and third is classification. The major disadvantage of this paper is that this system needs a large dataset.

3. CONCLUSION

This paper has evaluated various different types of research done in the field of Heart Failure prediction. The related works have given valuable insight into the various techniques that are being used to predict the onset of a Congestive Heart Failure. The authors have extensively utilized varied Machine Learning algorithms to accurately predict the signs of heart failure as it are one of the most crippling diseases that is usually fatal if not treated on time. The methodologies outlined in these works have provided an outline for an efficient and effective technique for the prediction of Congestive Heart Failure. After a thorough analysis of the machine learning prediction paradigm, we have designed an effective technique based on the Artificial Neural Networks and Dempster Shafer Reasoning which will be described in detail in the upcoming researches.

REFERENCES


****