International Journal of Medical Informatics 108 (2017) 134–145

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

A statistical analysis based recommender model for heart disease patients

Anam Mustaqeem[a], Syed Muhammad Anwar[a,*], Abdul Rashid Khan[b,c], Muhammad Majid[d]

[a] Department of Software Engineering, University of Engineering and Technology, Taxila, Pakistan
[b] Cardiology Department, Pakistan Ordinance Factories Hospital, Wah, Pakistan
[c] Wah Medical College, Wah, Pakistan
[d] Department of Computer Engineering, University of Engineering and Technology Taxila, Pakistan

ARTICLE INFO

Keywords:
Medical recommendations
E-health
Heart disease
Machine learning
Risk analysis

ABSTRACT

Objectives: An intelligent information technology based system could have a positive impact on the life-style of patients suffering from chronic diseases by providing useful health recommendations. In this paper, we have proposed a hybrid model that provides disease prediction and medical recommendations to cardiac patients. The first part aims at implementing a prediction model, that can identify the disease of a patient and classify it into one of the four output classes i.e., non-cardiac chest pain, silent ischemia, angina, and myocardial infarction. Following the disease prediction, the second part of the model provides general medical recommendations to patients.

Methods: The recommendations are generated by assessing the severity of clinical features of patients, estimating the risk associated with clinical features and disease, and calculating the probability of occurrence of disease. The purpose of this model is to build an intelligent and adaptive recommender system for heart disease patients. The experiments for the proposed recommender system are conducted on a clinical data set collected and labelled in consultation with medical experts from a known hospital.

Results: The performance of the proposed prediction model is evaluated using accuracy and kappa statistics as evaluation measures. The medical recommendations are generated based on information collected from a knowledge base created with the help of physicians. The results of the recommendation model are evaluated using confusion matrix and gives an accuracy of 97.8%.

Conclusion: The proposed system exhibits good prediction and recommendation accuracies and promises to be a useful contribution in the field of e-health and medical informatics.

1. Introduction

In recent times, most of our actions are monitored and stored in the form of data. It is now possible to predict the action or behavior of a person by extracting meaningful information from the historical data related to that person. Patients are also treated by medical advisors using the historical data, which accounts for different clinical factors. Healthcare systems organize data to maintain records for different diseases and take appropriate actions. The presence or absence of a disease can be predicted by monitoring certain clinical parameters. The real challenge is to assess the level of risk or severity, and then generate appropriate recommendations, either to cure the disease or to minimize the risks associated to that particular disease [1].

Chronic diseases are affected by day to day human life style, including diet and family history. One of the advanced topics of research, which has shown enormous amount of effort for providing an efficient response to chronic diseases is e-health. E-health based methods are applied to a wide variety of patients, which are remotely placed and are affected from several chronic diseases [1]. Among chronic diseases, cardiac disease is one of the major cause of death around the globe. Historically, a larger ratio of people in older age are found to be affected by cardiac diseases. During the past few years this ratio has significantly altered, whereby cardiac diseases are now widely observed in a younger age. Patients living in urban areas can easily get clinical treatment and protect their lives from threatening diseases. But people of remote areas are often seen struggling for their survival, due to limited availability of medical facilities. To reduce sickness, mortality, and expenses related to heart syndromes, e-health has emerged as a promising solution [2]. However, one of the main challenge faced by e-health solutions is the capability of gathering and processing information effectively and efficiently. A remote monitoring system is considered effective and robust, if it intelligently analyzes the data gathered from patients, and generate reports accordingly [3]. The historical clinical records are used for preventive actions, which can be taken to

* Corresponding author.
E-mail address: s.anwar@uettaxila.edu.pk (S.M. Anwar).

http://dx.doi.org/10.1016/j.ijmedinf.2017.10.008
Received 7 July 2017; Received in revised form 7 October 2017; Accepted 12 October 2017

reduce the impact of clinical disorders. Most of the existing remote health monitoring systems are deficient of such progressive analytical proficiencies [2,3].

In recent years, health monitoring systems have evolved to be used in a more challenging manner, rather than to be used in simpler ways, such as counting of sleep hours [4]. Health monitoring algorithms are used to process data in such a way, which is more useful in terms of the information provided to the end user [5]. In recent studies, advanced machine learning techniques have been used to represent more complex information in three main tasks i.e., (1) prediction, (2) anomaly detection, which can also be used for the alarm based machine learning algorithms, and (3) decision making process, which is used to classify input data in different classes already available for any particular disease [6,7]. However, it is still challenging to acquire data in an efficient and highly accurate analytic manner, which provides support in taking personalized decisions for diagnosis of a disease. The goal of e-health systems is to minimize the risk to human life by detecting various diseases at an earlier stage and provide relevant recommendations. A lot of work has been done to detect a particular disease at different clinical stages [7,8], but there is a significant need to provide appropriate and personalized medical recommendations to patients depending on their current clinical condition.

In this paper, a hybrid prediction and recommendation model is proposed for diagnosis and treatment of heart disease patients. The main aim is to propose an intelligent and adaptive recommender system for patients suffering from various heart diseases. Several prediction algorithms support the proposed recommender system to identify the disease and provide recommendations accordingly. Most of the work in this field has been done to identify any single heart disease and its sub types [9–11], whereas the proposed method provides a multi-classification approach to classify patients into one of the four major heart diseases i.e., non-cardiac chest pain (NCCP), silent ischemia (SI), angina, and myocardial infarction (MI). After assessing the results of the proposed prediction model, medical recommendations are provided to heart disease patients based on risk analysis and severity scores of the patient's medical data. The results of the proposed prediction system are evaluated using accuracy and kappa statistics as performance measures. Experimental estimations are conducted on labelled clinical data set of heart patients collected from a renowned local hospital. The recommendations are performed based on a knowledge base created with the help of medical experts, and risk analysis is performed to provide accurate recommendations to the patients. The results of the proposed system yield good prediction and recommendation accuracies, hence proving that the system will provide a significant contribution in heart disease prediction, risk assessment of various abnormalities associated with heart disease patients, and management of heart diseases. The proposed model has following contributions,

- A dataset is collected from a local hospital for heart disease patients, which is labelled by a team of medical experts for e-health recommendations.
- Feature selection and prediction is applied for the diagnosis of multiple heart diseases.
- The risk of predicted disease is assessed for generating relevant medical recommendations.
- Clinical experts verify the accuracy of the recommendation model.

The remaining paper consists of the following sections. In Section 2, related work in the field of classification and recommendations is presented. Section 3 describes an overview of the proposed system and is divided in two main parts. The first part explains the steps involved in disease classification and prediction, whereas the second part explains the methodology adopted for providing medical recommendations. The details of dataset and experimental results are discussed in Section 4. Finally, conclusions and future endeavors are highlighted in Section 5.

2. Related work

Cardiovascular diseases are one of the major cause of death in the last two decades [1]. The recent enhancements in computer technology and innovations in machine learning techniques have aided the development of efficient prediction and decision-making tools [8,12], which assist medical specialists in settling to an effective choice of treatment for various diseases at an early stage. The process of disease diagnosis relies heavily on pathological and clinical test data [13]. Data mining and machine learning algorithms are used for handling the issue of disease prediction using the available information. The coronary artery calcium score has been used to predict unknown coronary heart disease events [14]. It has been concluded, that the addition of coronary artery calcium score in traditional risk based prediction models, significantly improves the accuracy by placing more patients in the extreme risk class. Recently, artificial neural networks have gained a tremendous significance in the field of data mining and machine learning for classifying input data into various categories by discovering hidden dependencies in data, which proves to be helpful in predicting accurate classes [15].

One of the classification procedure used in e-health domain is the risk level prediction of disease in patients. Various machine learning algorithms, such as genetic algorithms, regression model, and decision trees have been successfully implemented to forecast the severity of a disease in patients [8,16–18]. The prediction models have been designed to evaluate the risk level of various diseases using clinical test readings, guidelines, signs, and symptoms. A suggested guideline of blood pressure, cholesterol, fasting blood sugar, and oxygen saturation effectively predicts the disease risk in various population samples [19]. A simple algorithm is presented in [20], which uses categorical attributes that assists physicians in the prediction of multivariate cardiac disease risk in patients. Also, generalized models and support vector machines (SVMs) have shown good performance in calculating risk for post-operative sepsis. Feature selection further improves the predictive accuracy of the proposed model [21]. However, by applying various techniques on heterogeneous patient data, it is concluded that machine learning based algorithms are more efficient in predicting the accurate cardiac disease, as compared to rule based risk prediction methods [15,22].

Classification of electrocardiogram (ECG) signals also contribute significantly in early diagnosis of various cardiac diseases. Feature extraction and classification are used sequentially to predict the occurrence of cardiovascular disease such as arrhythmia, which can be efficiently classified from heart beat rhythms [23,24]. Principal component analysis (PCA), fuzzy c-means (FCM), SVM and 1-D convolutional neural networks have been used for classification and extraction purposes [25–28] for diagnosis of cardiac patients. A 1-D convolutional neural network has proved to be a generic and accurate method in classifying the ECG records for any dataset [26].

Physicians usually classify patients into two categories i.e., either diseased or non-diseased, which are divided based on disease subtype or etiology. Decision tree algorithms are mostly used for classification purposes but it suffers from limited accuracy [8]. The utilization of tree based methods has been shown in [29], offering high performance compared to convolutional regression trees for classification of heart failure sub-types in patients from Canada. But, the improvements over logistic regression are not substantial. On the other hand, various machine learning algorithms such as Bayesian classifiers, logistic regression, decision trees, Naive bayes, j48 algorithm, bagging and back-propagation neural network have been used for early prediction of various diseases [9–11,30,31].

Intelligent methods and techniques also play a significant role in providing necessary medical recommendations to patients suffering from chronic diseases [32]. These useful recommendations can help in improving the life style of patients by reducing the workload and expenses involved in routine healthcare activities. In [33], a study has