INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 2 ISSUE 2 (2016) PAGES 847 - 852
RECEIVED : 27.03.2016 PUBLISHED : 05.04.2016
April 05, 2016

OVERVIEW OF HEART DISESE PREDICTION TECHNIQUES WITH THEIR FUNCTIONALITIES AND PERFORMANCE METRICS

[1]Purusothaman G, Ph.D, Research Scholar, Assistant Professor, School of Computer Studies (MCA) , RVS College of Arts Science, Sulur, Bharathiar University, Coimbatore, Tamil Nadu, India.

[2]Dr. A. Nithya, Assistant Professor, School of Computer Studies – Research and CS, RVS College of Arts Science, Sulur, Bharathiar University, Coimbatore, Tamil Nadu, India.

ABSTRACT

Heart disease is the most common issue found in peoples of all age categories. Heart disease would lead to various healths related threats which need to be focused well for the improved performance. Early prediction of heart disease would help peoples and doctors to treat them immediately which might avoid some serious threats. However early prediction of heart disease would be more difficult task due to non availability in necessary details. There are various research methods has been introduced earlier that focus on prediction of the heart disease in the accurate manner by following different strategies and methods. In this analysis work, the previous research methodologies have been discussed in detail which indicated their working procedure. This analysis work also shows overview about their working procedure along with merits and demerits of those research methodologies. This work provides the discussion about the heart disease prediction in sectionalized way. Those are novel feature selection techniques; risk factors identification and prediction procedures. These research methods are defined and indicated with their faults and advantages. The overall review of the research methods has been conducted in the matlab simulation environment from which it can be found that improvement of the research methodologies. This evaluation has been carried out based on merits and demerits of research methods that have been discussed.

Keywords: Heart disease prediction, feature selection, risk factor identification, functionality

1. INTRODUCTION

Heart disease found to be most occurred disease on people of all age groups for the past many years. It is found to be most major threat by the World Health Organization (WHO) [1]. It is confirmed that the major reason for the 41% of deaths happened in real world is the heart attacks, strokes and other circulatory diseases. This surveyed and proved by the European public health alliance [2]. Thus it is more essential to predict the occurrence of heart disease in the accurate manner with the consideration of the various security threats. Heart disease consists different kind of symptoms before it happens based on human health [3]. It makes more difficult to predict the heart disease accurately due to presence of various kind of symptoms. Manual involvement of doctors would make this process easier where they can manually assign weight values to the attributes based on their important level. This would make heart disease prediction process more flexible and convenient [4].

Medical experts for the heart disease handling are less in number in the real world where the manual prediction of heart disease would also be more complex [5]. Lesser medical experts would lead to less medical knowledge where there won't be sufficient information available for heart disease prediction. Thus it is required to implement the automated techniques which can predict the heart disease occurrence in the efficient way without any error. An decision making system can make it possible which can predict the

heart disease more accurately without human intervention. It required medical database to be processed in the accurate way [6]. The main contribution of this research work is to analyze and discuss about the various and different research methodologies that are conducted towards achieving the accurate heart disease prediction rate. These techniques have been discussed in detail in terms of their working procedure based on which comparison would be made. The merits and demerits of the research methodologies have been discussed in detail based on their working procedure based on which comparison would be made out. The overall organization of the research method is given as follows: In this section, general introduction about the heart disease prediction importance and characteristics has been given. In section 2, detailed review about the varying related research works that has been introduced by various researchers has been given. In section 3, comparison evaluation of the research methodologies has been given. Finally in section 4, conclusion of the analysis conducted in this research in terms of their merits and demerits has been given.

2. ANALYSIS OF RELATED RESEARCH WORKS

Heart disease is most common disease that is found in most of humans of all age groups. Early heart disease prediction needs to be done with more concern for the efficient treatment to be provided. There are various research techniques has been proposed earlier by the different researchers to predict the heart disease in the accurate manner. In this section heart disease prediction technique has been discussed in three categories. Those are feature selection, risk factor selection, classification. The detailed explanation of the heart disease prediction techniques has been given in the following sub sections.

FEATURE SELECTION TECHNIQUES USED FOR HEART DISEASE PREDICTION

This section provides the various feature selection techniques that help to improve the heart disease prediction rate has been discussed in detail. In [8], Ischemic Heart Disease Identification process is improvised by introducing the efficient feature selection techniques namely Multi Layer Perceptron Neural Network. This method adapts behaviour of Artificial Neural network to select the most optimal features from the set of attributes present in the heart disease database. The main goal of this research method is to select the most interesting and required features from the ischemic heart disease database which consists of records from the 712 patients. The evaluation procedure of this research work proves that the proposed research methods lead to provide the improved accuracy than the existing research methods. In [9], the performance cardiovascular disease prediction has been improved by focusing on the feature selection process. The main goal of this research method is, classification with interesting features instead of entire feature would increase the prediction rate. This method introduce the ranking based feature selection technique which is assured by utilizing the machine learning techniques such filter and wrapper methods. This method assures the efficient selection of feature subsets from the entire feature set. In [10], authors introduced the automated method for the important feature selection process. The main goal of this research method is to select the features which are disease specific one. This method performs both filtering and wrapping to reduce the feature dimensionalityIn the following table 1, content analysis of the above describe techniques has been given to know the merits and demerits of those techniques.

ANALYSIS OVERVIEW OF RISK FACTOR SELECTION TECHNIQUES

Risk factors are most prominent factors in the process of risk factor selection which needs to be focused well for the accurate and early prediction of heart disease. There are upto 300 risk factors have identified already which are reasonable for the coronary heart disease. Finding the presence of these risk factors would increase the prediction rate of heart disease. There are various research method has been introduced earlier by various researchers that focus on prediction of risk factors. Those research methods have been discussed here in this section. In [13], authors performed risk factor presence analysis in the

different kind of disease databases by adapting the non linear classifiers namely non linear support vector

Table 1. Analysis overview of feature selection techniques

S.No	Author and year	Method	Merits	Demerits
1	K.Rajeswari et al 2012	Multi Layer Perceptron Neural Network	Improved prediction accuracy with more training information	Reduction of features leads to reduced prediction accuracy rate
2	Swati Shilaskar, and Ashok Ghatol, 2013	Ranking based feature selection	Ensured selection of efficient feature subset.	More computation overhead
3	Zhang, Z et al 2014	one versus one feature selection and then the OvO-rule support vector machine (SVM) binary classifier	Increased prediction accuracy Improved sensitivity rate Increased classification accuracy	Requires more training information for the increased prediction accuracy rate
4	Schmidt, S. E et al 2015	a quadratic discriminant function	Improved prediction of coronary heart disease. Increased accuracy rate	Requires more information regarding detection of the features
5	Syed Muhammad Saqlain Shah et al 2017	Probabilistic Principal Component Analysis	Increased accuracy of prediction rate. More performance during statistical test report generation	Requires complete information in the feature subset to provide the accurate prediction result

Risk factors are most prominent factors in the process of risk factor selection which needs to be focused well for the accurate and early prediction of heart disease. There are upto 300 risk factors have identified already which are reasonable for the coronary heart disease. Finding the presence of these risk factors would increase the prediction rate of heart disease. There are various research method has been introduced earlier by various researchers that focus on prediction of risk factors. Those research methods have been discussed here in this section. In [13], authors performed risk factor presence analysis in the different kind of disease databases by adapting the non linear classifiers namely non linear support vector machine. The main goal of this research method is to search and predict the different kind of features present in the database. The performance of this process is improved by attempting different kernels namely nonlinear kernel, localized radial basis function kernel. The performance evaluation proves that the risk factors selected by using these techniques would increase the classification accuracy considerably. In [14], decision tree based risk factor analysis is done to accurately predict the presence of disease. This research work analysis two types of risk factors namely modifiable factors and non modifiable factors. The C4.5 algorithm has been adapted to perform the risk factor analysis which can select the most important features from the medical dataset more accurately. This method result with the more important factors such as smoking, hypertension and family history as risk factors which tend to lead into accurate classification rate.

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 2 ISSUE 2 (2016) PAGES 847 - 852
RECEIVED : 27.03.2016 PUBLISHED : 05.04.2016
April 05, 2016

Table 2. Analysis of risk factor selection techniques

S.No	Author and year	Method	Merits	Demerits
1	Baek Hwan Cho et al 2008	Nonlinear Support Vector Machine	RBF kernel performs better and select the more efficient risk factors. RBF can work better in case of presence of noises	LRBF kernel seems to degrade in its performance in case of presence of more noises.
2	Minas A. et al 2010	Decision tree	Accurate prediction of risk factors that are more reasonable for the heart disease occurrence	It cannot work better for the large databases which needs to be concerned better in future
3	Li-Na Pu et al 2012	Genetic information based prediction	Accurate prediction of risk factors. Can find the more variation between the risk factors that are associated	Requires more computation overhead which needs to be concentrated more for the accurate classification rate
4	Sree Hari Rao et al 2013	Particle swarm optimization approach	Increased accurate of risk factor selection. Can find better by making decision rule efficiently	Requires more computation time for generating and analyzing the decision rules
5	George Karystianis et al 2015	Local lexicalized rules	Increased F score value. Can identify the progression of heart disease even in case of presence of large volume of database	Require knowledge about the clinical note for the accurate selection of risk factors
6	Masaaki Matsunaga et al 2017	Association rule based similarity identification	It is found hyper tension as more important risk factor. Lead to optimal prediction of heart disease	Cannot perform well for large databases

ANALYSIS OF HEART DISEASE PREDICTION TECHNIQUES

In this section, various heart disease prediction techniques have been discussed in terms of their working procedure. The different metrics involved under those methodologies has been discussed in detail in this section. In [19], authors introduced statistical methods for the prediction of coronary heart disease occurrence on patients. This research work would be adjusting the risk factors that are reasonable for the heart disease occurrence by finding the discrimination difference between them. In [20], authors introduced firefly algorithm and the type 2 fuzzy logic system for the accurate prediction of heart disease.

This work predicts the heart disease by introducing the hybrid method where the integration of methods has been done which can select the more important features from the dataset. This method can assure the accurate prediction of heart disease with more computation overheads

Table 3. Prediction method comparison

S.No	Author and year	Method	Merits	Demerits
1	Hochholzer, W et al 2014	Statistical method	Accurate prediction of risk factors. Efficient heart disease prediction	Not efficient in case of patients with incomplete information which requires more time complexity to accomplish the task
2	Long, N. C et al 2015	Firefly based algorithm, interval type-2 fuzzy logic system	Increased accuracy rate. More computation efficiency	Increased expensive Need over concern
3	Saxena, K., & Sharma, R. 2016	Decision rule based system	Improved prediction rate. Higher classification accuracy	It might lead to worse result in case of insufficient training
4	Driscoll, A et al 2017	multivariable risk-factor model	Improved discrimination rate. Accurate heart disease prediction	Reclassification is required to provide the accurate classification rate

CONCLUSION

Heart disease prediction is the most concerned research issue focused by the various researchers to perform early prediction of heart disease. There is multiple research methods has been introduced earlier for prediction of heart disease in the accurate manner. In this research analysis paper, analysis multiple existing research methodologies has been discussed in terms of their working procedure and the performance metrics utilized. This analysis work also shows overview about their working procedure along with merits and demerits of those research methodologies. This work provides the discussion about the heart disease prediction in sectionalized way. Those are novel feature selection techniques; risk factors identification and prediction procedures. These research methods are defined and indicated with their faults and advantages. The overall review of the research methods has been conducted in the matlab simulation environment from which it can be found that improvement of the research methodologies. This evaluation has been carried out based on merits and demerits of research methods that have been discussed.

REFERENCE

1. World Health Organization. (2000). The world health report 2000: health systems: improving performance. World Health Organization.
2. Kickbusch, I. (2003). The contribution of the World Health Organization to a new public health and health promotion. American journal of public health, 93(3), 383-388.

3. MEMBERS, W. G., Go, A. S., Mozaffarian, D., Roger, V. L., Benjamin, E. J., Berry, J. D., ... & Franco, S. (2014). Heart disease and stroke statistics—2014 update: a report from the American Heart Association. circulation, 129(3), e28.
4. Jee, S. H., Jang, Y., Oh, D. J., Oh, B. H., Lee, S. H., Park, S. W., & Yun, Y. D. (2014). A coronary heart disease prediction model: the Korean Heart Study. BMJ open, 4(5), e005025.
5. Seera, M., & Lim, C. P. (2014). A hybrid intelligent system for medical data classification. *Expert Systems with Applications*, *41*(5), 2239-2249.
6. Kaur, B., & Singh, W. (2014). Review on heart disease prediction system using data mining techniques. *International journal on recent and innovation trends in computing and communication*, *2*(10), 3003-3008.
7. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
8. K.Rajeswari, Dr.V.Vaithiyanathan, Dr. T.R. Neelakantan, "Feature Selection in Ischemic Heart Disease Identification using Feed Forward Neural Networks", International Symposium on Robotics and Intelligent Sensors 2012, Procedia Engineering 41 (2012) 1818 – 1823
9. Swati Shilaskar, Ashok Ghatol, "Feature selection for medical diagnosis : Evaluation for cardiovascular diseases", Expert Systems with Applications 40 (2013) 4146–4153
10. Zhang, Z., Dong, J., Luo, X., Choi, K. S., & Wu, X. (2014). Heartbeat classification using disease-specific feature selection. *Computers in biology and medicine*, *46*, 79-89.
11. Schmidt, S. E., Holst-Hansen, C., Hansen, J., Toft, E., & Struijk, J. J. (2015). Acoustic features for the identification of coronary artery disease. *IEEE Transactions on Biomedical Engineering*, *62*(11), 2611-2619.
12. Syed Muhammad Saqlain Shah, Safeera Batool, Imran Khan, Muhammad Usman Ashrafa, Syed Hussnain Abbas, Syed Adnan Hussain, "Feature Extraction through Parallel Probabilistic Principal Component Analysis for Heart Disease Diagnosis", Physica A (2017), http://dx.doi.org/10.1016/j.physa.2017.04.113
13. Baek Hwan Cho, Hwanjo Yu, Jongshill Lee, Young Joon Chee, In Young Kim, and Sun I. Kim, "Nonlinear Support Vector Machine Visualization for Risk Factor Analysis Using Nomograms and Localized Radial Basis Function Kernels", IEEE Transactions On Information Technology In Biomedicine, Vol. 12, No. 2, March 2008
14. Minas A. Karaolis, Joseph A. Moutiris, Demetra Hadjipanayi, and Constantinos S. Pattichis, "Assessment of the Risk Factors of Coronary Heart Events Based on Data Mining With Decision Trees", IEEE Transactions On Information Technology In Biomedicine, Vol. 14, No. 3, May 2010
15. Li-Na Pu, Ze Zhao, and Yuan-Ting Zhang,, "Investigation on Cardiovascular Risk Prediction Using Genetic Information", IEEE Transactions On Information Technology In Biomedicine, Vol. 16, No. 5, September 2012
16. V. Sree Hari Rao, and M. Naresh Kumar, "Novel Approaches for Predicting Risk Factors of Atherosclerosis", IEEE Journal Of Biomedical And Health Informatics, Vol. 17, No. 1, January 2013
17. George Karystianis, Azad Dehghan, Aleksandar Kovacevic, John A. Keane, Goran Nenadic, "Using Local Lexicalized Rules to Identify Heart Disease Risk Factors in Clinical Notes", Journal of Biomedical Informatics (2015), doi: http://dx.doi.org/10.1016/j.jbi.2015.06.013