Survey on Clinical Decision Support System for Diagnosing Heart Disease

Suchithra¹, Dr. P. Uma Maheswari²

¹PG Scholar, Info Institute of Engineering, India
²Professor and Head, Department of CSE, Info Institute of Engineering, India

chanu.suchithra@gmail.com
dr.umasundar@gmail.com

Abstract— Ischemic Heart Disease is a disease which is difficult to diagnose and is very commonly identified only during the mortality of an individual. The World Health Organization (WHO)[12] statistical report state that more than 70 per cent of coronary deaths occur with subjects older than 70 years in North America and Western Europe. As per WHO reports in India and other developing countries 70 per cent deaths occur in subjects less than 70 years of age. Coronary Heart Disease (CHD) is an epidemic in India. A retrospective data set that included 1000 clinical cases is taken for the work. 88 sets were discarded during preprocessing. Tests were run on 912 cases using weka classifiers[5] available in weka 3.7.0. Out of 113 classifiers, 16 classifiers are identified to be the best based on different parameters sensitivity, specificity, accuracy, F-measure, kappa statistic, correctly classified cases, time taken to run the model and ROC curve. The diagnoses made by Clinical decision Support System(CDSS)[1][6][9] were compared with those made by physicians during patient consultations. The major goal of this paper is to build an expert system for diagnosing the presence of Ischemic Heart Disease with an integrated automated classifier using Artificial Intelligence techniques.

Keywords— Artificial Intelligence Techniques; Clinical Decision Support System (CDSS); Ischemic Heart Disease; Kappa Statistics; Sensitivity; Specificity

I. INTRODUCTION

At present India is undergoing an epidemiological transition and one of the threshold is the epidemic cardiovascular disease. The statistical cause-specific mortality rate data indicate that cardiovascular disease plays an important role in contribution to mortality. Demographic projections graph suggest a major increase in cardiovascular disease as the growing population changes in rate with morality and life. Surveys in urban areas provides information that coronary risk factors are already widespread and that spontaneous actions are needed to be handled for preventing a further rise as socioeconomic development proceeds. It is important to obtain epidemiological data from various regions in order to plan, initiate and monitor public health action.

This work focuses on different classification algorithms applied on a dataset collected from Madras Medical College. A comparison of different measures like sensitivity, specificity kappa statistic, ROC[9], time taken for classification is made. Limitations like all patients in the heart disease dataset has to receive three non-invasive cardiac tests and complete clinical and ECG data to be collected from all the patients are reduced to a greater extent using the native dataset and different artificial intelligence techniques.
Diagnosis criteria in medicine at western regions are done by referring to —Naming and diagnosis criteria of ischemic heart disease which was issued by International Society of Cardiology and the Joint Subject Team on standardization of clinical naming in World Health Organization. At Present India is undergoing a rapid health transition with rising spread of CHD.

This work undergoes with the classification of risk level for heart disease that might cause sudden death. Knowing about the risk level, people can take the required adequate measures depending on the severity of the disease or follow very strict dietary rules to avoid future problems that might cause mortality.

Artificial intelligence (AI)[1][2][10][24][32] is a branch of computer science that includes study and development of intelligent machines and software. Major AI researchers and textbooks define this field as "the study and design of intelligent agents", where an intelligent agent is a system that perceives its environment and takes actions that maximize its chances of success.

Cardiovascular Diseases in India

Fig 1.1 Percentage of Mortality

Fig 1.2 Major causes of death in India Rural Vs Urban

Fig 1.3 A wide spread CVD in India - Article
Table 1.1 Top 10 causes of Morality

<table>
<thead>
<tr>
<th>2008</th>
<th>2030</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiovascular diseases</td>
<td>Cardiovascular diseases</td>
</tr>
<tr>
<td>Cancers</td>
<td>Cancers</td>
</tr>
<tr>
<td>Chronic Respiratory</td>
<td>Chronic Respiratory</td>
</tr>
<tr>
<td>diseases</td>
<td>diseases</td>
</tr>
<tr>
<td>Respiratory infections</td>
<td>Respiratory infections</td>
</tr>
<tr>
<td>Perinatal Conditions</td>
<td>Diabetes Mellitus</td>
</tr>
<tr>
<td>Diarrhoeal diseases</td>
<td>Digestive diseases</td>
</tr>
<tr>
<td>Digestive diseases</td>
<td>Perinatal Conditions</td>
</tr>
<tr>
<td>HIV/AIDS</td>
<td>Neuropsychiatric disorders</td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>Genitourinary diseases</td>
</tr>
<tr>
<td>Neuropsychiatric conditions</td>
<td>HIV/AIDS</td>
</tr>
</tbody>
</table>

II. LITERATURE SURVEY

The following have been analysed and studied in order to develop an alert system for heart disease in India.

The World Health Statistics released in 2012 by the World Health Organization (WHO) clearly describes the worrying trends for Indians. Among adults:
- Over 20 years of age - the estimated prevalence of CHD is: 3-4% in rural areas & 8-10% in urban areas, representing a two-fold rise in rural areas and a six-fold rise in urban areas between the years 1960 and 2000.
- In comparison with the population in the United States, the presence of CHD in Asian Indians is approximately 4 times higher.

According to WHO statistics, as many as 24% men and 22.6% women in India - aged 25 years and above - are suffering from high blood pressure (BP).

As per Statistical Reports:
- One in 10 men and women aged 25 years and above have high sugar.
- More women in India (2.5% of adults aged 20 years and above) are obese compared to Indian men (1.3%).
- Almost one in five (19%) boys, aged 13-15 years (adolescents), and 8% girls smoke tobacco.

European Union experts say that there still exist drastic change in disparities between countries not only in terms of CVD incidence, but also with regard to national prevention policies. According to them, obese adolescents have no symptoms of heart disease or any damaged hearts with thicker walls. Moreover, both structural and functional measures correlate with Body Mass Index (BMI). That is why these findings have explicate reason why obesity is a risk for heart disease. Therefore, long-term case fatality following acute coronary syndrome is considerably higher among Indians as compared to other country populations. In addition to this, there are several reversal of socio-economic gradients for the CHD factors that emerged in the Indian Population.

The problem that persists in the present situation is that it's of less expert doctors plays the major impact for the motivation of our problem statement.

The Statistical Report indicates the following:
- As per the Medical Council of India, the total number of registered allopathic doctors in the country is as little as under 7 lakh. Current population of India is 1.22 Billion.
- About 72.2% of the population lives in some 638,000 villages and the rest 27.8% in about 5,480 towns and urban agglomerations.
- According to this statistic, only one allopathic doctor is available for over 1,600 patients.
- As per the medical council the total allopathic doctors produced per year are 12,000.
- This brings a miserable situation of measure as just 1 doctor to a population of 2,000.
- And the population of nurses as, it is just 1 for every 2,200 patients.

2.1 Artificial Intelligence techniques applied to the development of a clinical decision support system for diagnosing Ischemic Heart Disease.

Dr. Vaidyanathan and K. Rajeswari et al., 2012 stated in their work as: A heart system inquiry diagnosis scale is being designed, where the symptoms are defined clearly, and therefore the detailed collecting methods are listed which helps in easy identification of the CVD. The total dataset is divided into training and testing data. A 10-fold cross validation is being used. The different classifiers of Weka are trained with the training set and therefore the sensitivity and the specificity of the classifier are evaluated using the test set. This method helps in reinforcing the validation process, such that the sensitivity and specificity values are the average of the ten validation folds. A training set is a set of data used for discovering potentially related predictive relationship between the given input and the produced output. This study was approved by Madras Medical College and SASTRA University and was conducted in Cardio Thoracic Department of Madras Medical College.

2.2 International Heart Protection Summit September 2011.

KS Reddy et al submitted a statistics in 2011 as: At present India is undergoing an epidemiological transition on the threshold of an epidemic cardiovascular disease. This paper suggests the need of Information Technology and its implementation towards the medical field. According to this paper, a Decision Support System is indeed needed in order to control the wide spread epidemic CardioVascular Disease. The system is being designed for Indian Population. The suggestion of this paper is that Coronary Heart Disease or Ischemic Heart Disease can be handled successfully if more research is encouraged in this mixed medical cum technology area.

2.3 Classification of Risk Level for Ischemic Heart Disease in India using Artificial Intelligence.

Dr. P. Amirtharaj et al 2011 discussed Cardiovascular Diseases (CVD) contains of a group of diseases that concerns with the heart and vascular system. The major condition of this disease include Coronary Heart Disease (CHD) or Ischemic Heart Disease (IHD) which causes 25-30 percent of deaths (major part in death percentage) in most of the industrialized countries. India too comes under this and is in a risk of developing more deaths due to CHD. Therfore a Decision Support System (DSS)[11] is being proposed in order to identify the level of risk for Ischemic Heart Disease of a Patient. This would help patients in taking the required precautionary steps like: following a balanced diet and medication which in turn may increase the life time of a patient. The attributes for prediction are being selected under certain factors i.e., after considering Indian conditions from literature and on the Expert advice from Doctors. Framingham Risk score is used in this paper which has five attributes is for comparison. This system might change and might have fourteen features to be analyzed according to Indian Conditions. The system is a theoretical study which proposes implementation of Artificial Intelligence to mine the knowledge from Medical data. This paper concludes that by the end of next year, India will have 60% of the world's heart disease burden. The statistical report says that when compared to people in other developed countries, the average age of patients with heart disease is at least 5-8 years lower among Indians than from the western areas. Normally, sixty is the average age of heart patients in India against 63-68 in developed countries. Now at this scenario the age is slipping further to the mid-50s. Indians are more likely to have types of heart disease like ischemic heart disease — a condition characterized by reduced blood supply to the heart.

2.4 Prediction of Risk Score for Heart Disease in India using Machine Intelligence.

Dr. P. Amirrharaj et al 2012 concentrated on the following: Machine Intelligence is an area used for Medical Data Mining. Medical Data Mining by term defines in finding interesting information from a large collection of Medical data. This paper helps us to find useful information from heart disease data set. This paper helps in theoretical study in the implementation of Machine Intelligence algorithms. This paper states that the system has nineteen features. To reduce the number of features Genetic Algorithm is being used. The importance of data mining techniques is very paly a role in providing better patient care and effective diagnostic capabilities by finding exact patterns and extracting knowledge which increases with the increase in the volume of stored data.

2.5 Cardiovascular Diseases & its Impact – WHO

WHO 2010 addresses the following: Coronary risk factors are widespread and that a disease which needs urgent action to prevent a further rise in socioeconomic development proceeds. A Clinical Decision Support System (CDSS) for reliable heart disease risk classification using Artificial Intelligent techniques have been proposed. As this approach focuses on CAD Risk analysis, for a sample population, future work may be directed for further analysis.
National Status:
- In recent times, though ad-hoc disease surveys, a systematic approach is used for tracking diseases which does not exist.
- According to the publication of IJMR, it has been concluded that at national level critical components standard disease management guidelines, diet and physical activity guidelines are still not developed with intelligent computational assistance.

International Status:
- Substantial research work is in progress all over the world especially in countries like USA, UK, Europe, Japan and China. Despite no successful implementations still published. Even so, their products are much expensive and unaffordable by the public of India like countries.

2.6 Burden of Cardiovascular Diseases in India
By Rajeev Gupta & his associates 2007 the following were projected: The population of India is about more than 30 million people with diabetes. The numbers of diabetics climb steeply day by day. It is estimated that by 2025, India will be number one in the world with the maximum number of people with diabetes. This is according to the World Health Organization (WHO)'s press release in 2012. According to this publication by WHO, those people who are affected with diabetes will be in the age group 40 and 64. With the spread of fast-food outlets and more sedentary lifestyles, the prevalence of diabetes in India is rising alarmingly. It is a need of an expert system at the end to prevent the on-growth of the life-demanding Coronary Heart Disease.

2.7 Modelling Effective Diagnosis of Risk Complications in Type 2 Diabetes – A Predictive model for Indian Situation
T. Gurumoorthy et al. 2012 found a solution as: Diabetes Mellitus, also called as High Sugar Problem is a metabolic disorder described by chronic hyperglycemia enhancing rapidly. It is a polygenic disease which is characterized by abnormal high glucose in the blood flow. Statistics say that 90 to 95% of the World Diabetics have Type 2 Diabetes. And India tops the list above all countries. The consequences of diabetes may be macro vascular complication or micro vascular complication. A new model for prediction of complications developing due to Diabetes Mellitus is proposed. Data collection is done after discussion with Diabetologists of Diabetes care and Research centre. A questionnaire is prepared with seven stages after getting concern from Expert Doctor’s opinion and Artificial Neural Network technique is used to predict the complications developing. As many researchers have contributed in diagnosis of Type 2 Diabetes, our work focuses on modelling an effective Diagnosis of a special complication called neuropathy.

2.8A Novel Risk Level Classification of Ischemic Heart Disease using Artificial Neural Network Technique – An Indian Case Study
Dr. V Vaithiyanathan and Dr. P. Amirtharaj et al. 2012 found the following: A Decision Support System (DSS) is proposed is highly needed to identify the level of risk in Ischemic Heart Disease for a Patient. Framingham Risk score includes five attributes which is used for comparison. The proposed system consists of seventeen features which has to be analyzed accordingly to the Indian Conditions. The system proposes the implementation of Artificial Neural Network technique to mine the knowledge from Medical data collected.

2.9 Heart disease diagnosis: an efficient decision support system based on fuzzy logic and genetic algorithm.
K. Rajeswari & V. Vaithiyanathan 2010 et al. stated the following as: Computerised clinical guidelines provide benefit for health outcomes and costs; however, the effective implementation has much more significant problems. One effective solution is to achieve an optimal trade-off between data ambiguity and good decision-making which would be to integrated data mining and artificial intelligence techniques. An efficient clinical decision support system (CDSS) for heart disease diagnosis using data mining and AI techniques is devised, which is also termed as an Expert System. The proposed algorithm makes use of the association pattern mining algorithm, apriori and genetic algorithm (GA) to formalise decision support architecture. We develop a clinical decision support system (CDSS) for diagnosis of IHD using Data Mining and AI techniques. The proposed expert system utilises Association pattern mining[23], GA[20] and fuzzy logic (FL) [31]for effective disease diagnosis. The input to the system is a clinical dataset that contains the medical records of heart patients which are usually the major attributes that causes CVD. The dataset is being preprocessed to eliminate noisy and insignificant information, such as the null values, irrelevant data (data entry errors, out of range data) and more from the data set, to prepare it adequately for the mining process. Results are being generated from the classified dataset.

2.10 Image Enhancement Techniques for Improving the Quality of Colour and Gray scale Medical Images.
The key objective presented by Preethi S.J. and K. Rajeswari 2009 is to process a medical image such that the results will be more suitable than the original image for a medical application, which can be achieved only by applying range compression, contrast stretching, histogram equalization, noise smoothing algorithms. Digital Medical Images[22] are usually
affected by unwanted noise, blurriness and also suffer from lack of contrast and sharpness which sometimes results in false diagnosis. This paper concentrates on eliminating those problems there by making the diagnosis easy.

2.11 Fuzzy based modeling for diabetic diagnostic decision support using Artificial Neural Network

K. Rajeswari & V. Vaithiyanathan 2011 investigates a variation from preliminary inquiry information that are obtained from the patients of diabetic and research center using fuzzy relation based model. The proposed model in this paper is an attempt closely to replicate a physician's insight of symptom-disease association and his approximate-reasoning for conclusion. The algorithm used here evaluated on a dataset of 1600 cases. The study is mainly on people who are approaching diabetician with either past history of Diabeties or with a new case of or with a symptoms of diabetics. Some cases are Normal patients without diabetics. The required attributes are estimated by interviewing patients. Later the attributes are modelled using a fuzzy approach which is being normalized and classified by Artificial neural networks as ‘Close to Type 2 diabetic’[24] or not. This result indicate the effectiveness of the proposed algorithm to an optimal model of the diagnosis for small or large datasets.

III. PROPOSED SYSTEM

The data warehouse of heart disease contains the screening data of heart disorder patients under evaluation. Initially, this data warehouse is being pre-processed either manually or using data mining tools to make the mining process more efficient. The algorithms are applied directly to the dataset. The data are being classified using the data mining tool. Usually WEKA[6][8] data mining tool is widely used for classification. The dataset is generally split into training and test data. 10-fold cross validation is used in this work, where the process is out of 10 tuples, 9 are taken as training samples and 1 as testing sample. Evaluation is usually described in terms of accuracy. The main goal of different supervised machine learning algorithms is accuracy improvement. As medical knowledge is vague, fuzzy sets are used to deal with uncertain linguistic medical concepts such as Less, Very Less, Medium.

The Overall Architecture of the proposed system is as follows:

The processing view of the system is:
Advantages:
- Helps in earlier identification of heart diseases.
- Reduction of death rate due to cardiovascular disease

IV. CONCLUSIONS

Based on the results obtained from the above, the following conclusions were made:
This study helps people in determining their heart disease risk, as it involves a simple procedure for decision making and taking
in the most effective way extracting hidden knowledge from the historical database which even might be a knowledge base.A
set of attributes relating to signs and symptoms and high-risk groups is being proposed that has been tested and has been proved
to be proficient and efficient in recognizing cases of IHD. The preliminary results of this work suggest that CVD is being
diagnosed using clinical decision support systems. KSTAR[17][19] algorithm helps in segregation of the negative and the
positive cases more appropriately than the other methods. Sensitivity and Accuracy are much better than the other methods. It
is more essential to have future future studies on classifier accuracy and its working proficiency including attribute selection for
IHD in developing an electronic protocol for CDSS.

For tracking risk factors population based adult cohorts has been set up:
- Identification of determinants regarding risk factors using cross sectional and prospective designs.
- Associating risk factors with CVDs using prospective design. (e.g., PURE study).
- Case-control study design which helps in rapid assessment of risk factors under adequate funding and supervision.
- Risk factor biology and genomics.

REFERENCES
[1] Josceli Maria Tenório, Anderson Diniz Hummel, Frederico Molina Cohrs, Vera Luci Sdeplanian, Ivan Torres Pisa, 
Heimar de Fátima Marin, —Artificial intelligence techniques applied to the development of a decision—support
cardiovascular risk factors to familial premature coronary artery disease: the GENE CARD project. J Am Coll Cardiol 
2002.
Intelligence in Medicine Vol.50., 2010.
Chinese Integrative Medicine, Vol.7PP:1., 2009.
[6] International Society of Cardiology and the Joint Subject Group on standardization of clinical naming in World Health 
Organization: Naming and diagnosis criteria of ischemic heart disease. Circulation Vol. 59 PP:3., 2012.44
SIGKDD Explorations 11 (1)(2009).
San Francisco, 2005.
[18] Newman, D.J., Hettich, S., Blake, C.L., Merz, C.J., UCI repository of machine learning databases. Department of 
[20] E. Massad, A teoriabayesiana no diagnósticomédico, in: E Massad, RX Menezes, PSP Silveira, NRS. Ortega (Eds.),
[22] P.K. Anooj, _Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules_,
Application to Cardiovascular Diseases_, IEEE Transactions on BioMedical Engineering,
[26] Ali Gharaviri, Member, Mohammad Teshnehlab and H. A. Moghaddam, _Ischemia Detection via ECG Using ANFIS_,
978-1-4244-1815-2/08/$25.00 ©2008 IEEE.46
[27] Minas A. Karaolis, Member, Joseph A. Moutiris, Demetra Hadjiapani, and Constantinos S. Pattichis, _Assessment of
the Risk Factors of Coronary Heart Events Based on Data MiningWith Decision Trees_, IEEE Transactions on
[28] D Frenkel, J Nadal, _Ischemic Episode Detection using an Artificial Neural Network Trained with Isolated ST-T
Segments_, 0276-6547/99 $10.00 © 1999 IEEE.
[29] Minghao Piao, Heon Gyu Lee, Guo Yong Sohn, Gouchol Pok, Keun Ho Ryo, _EmergingPatterns based Methodology for
Prediction of Patients with Myocardial Ischemia_, 978-0- 7695-3735-1/09 $25.00 © 2009 IEEE.
[30] National Cardiovascular Disease Database, Sticker No: SE / 04 / 233208,IC Health, Supported by Ministry of Health
[32] K Rajeswari, Dr.V. Vaithiyananathan and Dr.P. Amirtharaj, _Classification of Risk Level For Ischemic Heart Disease in
India using ArtificialIntelligence‘, ICMLC 3rd International Conference on MachineLearning and Computing vol5
Page 201-204., 2011.