Literature Survey on Investigation of Chronic Disease Correlation utilizing Data Mining Techniques

Babitha M
PG Scholar
Department of Computer Science and Engineering
Adhiyamaan College of Engineering, India

Manikandan M
Assistant Professor
Department of Computer Science and Engineering
Adhiyamaan College of Engineering, India

Abstract

People in today's world get affected by many diseases that do not have a complete cure. The development of one disease may lead to various other complications. A disease is an abnormal condition that affects the structure and function of one or more parts of the body. It may be caused by various factors, external and internal dysfunctions. There is a trend of various chronic diseases in any society. The major concern is that these chronic diseases are leading to many other diseases in future. An attempt to explore the correlation of various chronic diseases has become a necessity. This can be achieved by using data mining techniques, which help to derive knowledge about the affects of a particular chronic disease on the other chronic diseases. Since there is growing trend of diabetes and ischemic heart disease in the society, in this paper the focus is to investigate the effect of these diseases on the other chronic diseases using the ICD9 diagnostic codes. To achieve this goal various types of data mining techniques are used. The large amounts of data are very important in the field of data mining to extract useful information and generate relationships amongst the attributes. The conclusion is an optimal set of ICD9 diagnostic codes associated with individuals having diabetes or ischemic heart disease. These codes are then examined in the light of the human anatomic systems i.e. Circulatory system, Respiratory system, Nervous system, Musculoskeletal system, Renal system and Neoplasm and their relevance is justified.
Keywords: Chronic disease; ICD9 diagnostic codes; datamining; human anatomic system

I. INTRODUCTION

Illness is regularly connected with side effects and sign. These indications and sign describe an illness. One such category of a disease is chronic disease, which lasts for a longer duration. There are many chronic diseases like diabetes, ischemic heart disease, hypertension, kidney disease etc. which have to be controlled using medications and lifestyle management. In this course of span, these ceaseless illnesses may prompt numerous different infections. For instance, hypertension over a timeframe prompts coronary illness. Also diabetes too prompts numerous other therapeutic conditions later on stages. This might be because of a few reasons like determined utilization of medications, which influences the usefulness of different organs.

There is a need to understand the diseases and medical conditions which are caused in future as a result of chronic diseases. This learning helps the medical practitioner and the individuals to take precautionary measures. As mentioned there are many chronic diseases, but over the recent years there is growing trend of diabetes and ischemic heart disease in the urban population [1]. There are lot of medical conditions observed in individuals because of diabetes and ischemic heart disease. This knowledge can be derived by exploring the ICD9 diagnostic codes for the combination of each of these diseases with the other chronic diseases. Data mining techniques help to derive this knowledge. There are different categories in data mining techniques that can be applied to the healthcare data.

This paper focuses on the need to understand the diseases, medical conditions that occur in the near future if an individual has diabetes or ischemic heart disease. Experiments are performed utilizing the CMS (Center for Medicare Services) repository. Various data mining techniques like attribute reduction, resampling, classification are used. The data is extracted based on diabetes and ischemic heart diseases and its corresponding diagnostic codes. The reduced set of diagnostic codes obtained after applying the data mining techniques are explored based on the various anatomic systems of the human body namely circulatory system, nervous system, respiratory system, musculoskeletal system, renal system and neoplasm. The impact of diabetes and ischemic coronary illness on each of these frameworks is explored and successful conclusions are made.

II. LITERATURE SURVEY

Table - 1

#	Title	Year	Author	Methodology	Advantages	Disadvantages
1	A NEURAL NETWORK APPROACH FOR CLASSIFICATION OF	2015	Sergio Di Bona, Gabriele Pieri, Ovidio Salvetti	Error Back propagation algorithm:	The training function updates weights according to resilient	This behaviour is fully understandable, and it is due to

	KIDNEY DISEASE DATASETS				back-propagation algorithm.	the fact that quite often it is less expensive to reject.
					The low number of inputs and the lack of hidden neurons allow the algorithm to be trainable in a very short time	A single voxel and optionally leave a human to check it, than to accept wrong classifications
2	Review of Heart Disease Prediction System Using Data Mining and Hybrid Intelligent Techniques	2014	R.Chitraand V.Seenivasagam	CRISP-DM methodology	This list may need to be expanded to provide a more comprehensive diagnosis system.	Access to detailed information and accuracy.
					The system extracts hidden knowledge from a historical heart disease database.	Although notthe most effective model, Decision Trees results areeasier to read and interpret
3	Early Prediction of Heart Diseases Using Data Mining Techniques	2012	Vikas Chaurasia, Saurabh Pal	Decision Tree Classification	While claims data have previously been used to predict future health outcomes of patients.	The utility of the model had been recognized due to its ability to predict "less obvious" patients who are at risk for a preventable event.
					The dictates of cost efficiency behooved us to seek out higher precision in classification.	
4	Predictive Data Mining For Medical Diagnosis: An Overview Of Heart Disease Prediction	2012	Jyoti Soni, Ujma AnsariDipesh Sharma	FP-Growth Algorithm	The common symptoms of the diseases in a cluster can be used to describe or predict that group of diseases.	This contributes into producingsmaller and less complicated network and enhancinggeneralization ability on presented data due to the removableof redundant variables..
					Classification rule mining then extracts a small set of these rules and use them to build a classifier. To enhance the efficiency of discovering	
5	DETECTION OF LEUKEMIA	2015	Shubhangi Khobragade	leukemia detection	Hence automatic technique is adopted for fast and accurate results..	Pathological tests are costly and timely, so this process is not use for every month or short duration.
6	Assessment of Pulse Wave Velocity and Augmentation Index in different arteries in patients with severe coronary heart disease	2007	Irina Hlimonenko	Coronary artery calcification	The advantage of being non-invasive and do not require the administration of nephro-toxic radio-contrast agent.	Inadequate sample size and uncontrolled confounding are possible limitations, but are unlikely to have changed the main study findings.
7	Towards higher accuracy in supervised learning and dimensionality reduction by attribute subset selection - A pragmatic analysis	2012	D. Asir Antony Gnana Singh	J48 Classifier	The modified J48 classifier has been used to increase the accuracy rate of the data mining procedure.	Even the process is more flexible, it takes more time to classify due to the regression of decision tree.
8	A Distributed Approach to Alarm Management in Chronic Kidney Disease	2014	Miguel A. Estudillo-Valderrama	RRD(renal replacement therapy)	where there is less availability of and access to dialysis services, and is	It requires more number of datasets

					therefore an underestimate of the true demand.	
9	*Role of attributes selection in classification of Chronic Kidney Disease patients*	*2015*	*Naganna Chetty*	*K-means algorithms*	*An efficient model to predict and classify the CKD patient in the initial stage of CKD.*	*The models have shown better performance in classifying CKD and non CKD cases*
10	*Stage diagnosis for Chronic Kidney Disease based on ultrasonography*	*2014*	*Chi-Jim Chen*	*glomerular filtration rates (GFR).*	*Diagnosis through non-invasive ultrasonic imaging techniques become important clinical approaches for detecting CKD*	*This platform does not anyhow govern the input or output mappings of UIs*

III. PROPOSED DESIGN

After studying the exhaustive literature work on the research done, a design has been proposed. The various stages of the approach has been explained and depicted in the Figure I and explained in the following sections.

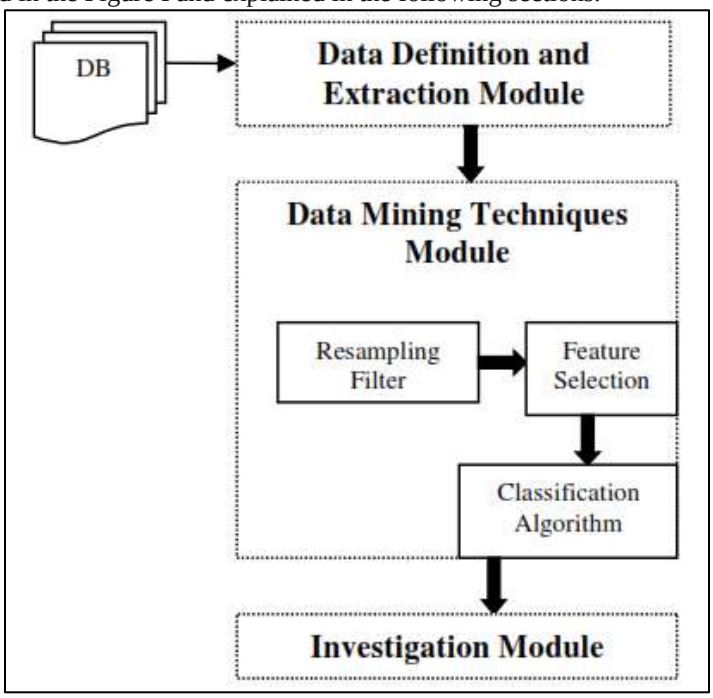

Fig. 1: Proposed Design

A. Data Definition and Extraction Module:

The information from the database (DB) is rebuilt i.e. the Beneficiary, Inpatient files and Outpatient files are connected using the primary key i.e. beneficiary id. The ICD9 diagnostic codes for the IP and OP files are also included in the data definition along with the chronic diseases. This definition helps to explore the ICD9 diagnostic codes and the chronic diseases effectively. Then the data is extracted by selecting the unique 2 digit and 3 digit ICD9 diagnostic codes, and the top N among them is extracted based on their occurrence. At that point division of information into tests in view of the interminable infections consolidated with each of the other endless ailments is finished. Here other chronic diseases along with diabetes and ischemic heart disease are extracted e.g. beneficiaries having diabetes and arthritis. Similarly every blend of 10 chronic disease with diabetes and ischemic heart disease is extracted. If a beneficiary has a particular chronic disease combination a value 1 is assigned this forms class C1 and value 2 for healthy patients which gives class C2.

B. Data Mining Techniques Module:

The next step is to apply various data mining techniques on the extracted data. It was observed that in the extracted data number of healthy individuals outnumber the beneficiaries with chronic diseases. This is a common observation in most of the healthcare datasets. This is leads to bias to a particular class during the learning process. To avoid this, the extracted data is resampled.

Resampling is one of the information mining procedure which guarantees uniform class dispersion, it appropriates the dataset consistently. At that point highlight choice systems are utilized to get the diminished arrangement of ICD9 codes. At that point these decreased arrangement of analytic codes are accepted utilizing an order calculation. Cross acceptance system is utilized for testing the decreased arrangement of codes.

C. *Investigation Module:*

Investigation is done by understanding the effect of diabetes and ischemic heart disease on various human anatomy system based on the ICD9 codes obtained. The anatomic frameworks researched are Circulatory system which incorporates illnesses identified with the heart i.e. Ischemic coronary illness and Chronic Heart disappointment, Nervous system which comprises of maladies identified with the cerebrum i.e. Stroke, Alzheimer, Depression, then Renal System that have disarranges identified with the kidney, Respiratory framework which is identified with the human breath i.e. fanatical pneumonic issue, Musculoskeletal framework that comprises of sicknesses identified with the bone and muscle in particular osteoporosis and joint inflammation lastly Neoplasm which influences the human invulnerability which incorporates malignancy. These diminished arrangement of symptomatic codes are cross approved and dissected by utilizing precision as the execution metric. Examination on the clinical significance of these diminished arrangement of ICD9 codes is done.

IV. CONCLUSION

The beneficiary file, , IP record, OP document from the CMS data set was restructured and relevant information is extracted based on the combination of chronic diseases i.e. for diabetes and ischemic heart disease. To achieve this goal various types of data mining techniques are used. The conclusion is an ideal arrangement of ICD9 demonstrative codes connected with people having diabetes or ischemic coronary illness.. These codes are then investigated based on the human anatomic systems i.e. Circulatory system, Respiratory system, Nervous system, musculoskeletal system, renal system and Neoplasm and their relevance is justified.

REFERENCES

[1] FrankB.Hu, Available from: http://care.diabetesjournals.org/ June 2011
[2] Peter B. Jensen, Lars J. Jensen, Søren Brunak,"Mining electronic health records: towards better research applications and clinical care",Nature Reviews Genetics 2012, 13:395-405.
[3] S.Priya, R.R.Rajalaxmi, "An Improved Data Mining Model to Predict the Occurrence of Type-2 Diabetes using Neural Network ", IJCA Proceedings on International Conference in Recent trends in Computational Methods,Communication and Controls, April 2012, pp:26-30.
[4] Diabetes mellitus: http://en.wikipedia.org/wiki/Diabetes_mellitus
[5] Advanced Vision Care: http://www.advancedvisioncare.com/conditions.php
[6] K.R.Lakshmi,Y.Nagesh,M.VeeraKrishna, "Performance Comparison of three data mining techniques for predicting kidney dialysis survivalility", International Journal of Advances in Engineering &Technology, Mar 2014, Vol.7, Issue 1, pp.242-254.
[7] Shruti Ratnakar, K. Rajeshwari, Rose Jacob, "Prediction of Heart Disease Using Genetic Algorithm for Selection of Optimal Reduced Set of Attributes", International Journal of Advanced Computational Engineering and Networking, April 2013, Volume-1, Issue-2, 2320- 2106.
[8] Nidhi Bhatla, Kiran Jyoti "A Novel Approach for Heart Disease Diagnostic using Data Mining and Fuzzy Logic". International Journal of Computer Applications, September 2012, Volume 54– No.17, pp:16- 21.
[9] GurbuzE, Biligisayar Muhendisilgi, Ondokuz, "Diagnostic of diabetes using Adaptive SVM and feature selection", Signal Processing and Communications Applications (SIU), IEEE 19th conference, April 2011, pp: 42-45.
[10] Alp Aslandogan et. al.," Evidence Combination in Medical Data Mining", Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04) 0-7695-2108-8/04©2004 IEEE.
[11] Sellappan Palaniappan, Rafiah Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", (IJCSNS), Vol.8 No.8, August 2008.
[12] Boleslaw Szymanski, Long Han, Mark Embrechts, Alexander Ross, Karsten Sternickel,Lijuan Zhu, "Using Efficient Supanova Kernel For Heart Disease Diagnosis", proc. ANNIE 06, intelligent engineering systems through artificial neural networks, vol. 16,page no. 305-310
[13] R.Sivakumar, G.Ravindran, M.Muthayya, S.Lakshminarayanan, C.U. Velmurughendran, ‖Diabetic Retinopathy Classification‖, IEEE, 2003.
[14] R.Sivakumar, G.Ravindran, M.Muthayya, S.Lakshminarayanan, C.U. Velmurughendran, ‖Diabetic Retinopathy Classification‖, IEEE, 2003.
[15] Asha Rajkumar and B. Sophia Reena, "Diagnosis Of Heart Disease Using Data mining Algorithm" , Global Journal of Computer Science and Technology, Vol. 10, No. 10, pp. 38 - 43, 2010