International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 12, December 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

CLINICAL HEALTH RISK PREDICTION OF CHRONIC DISEASES FROM SEQUENTIAL PATTERN MINING

Mr jaya prakash k[1], Mr. M. Chandrakumar[2]

[1] MCA., M.PHIL SCHOLAR,

[2]MCA., M.PHIL., PGDCA., PGDMM.,

HOD OF BCA, SNMV COLLEGE OF ARTS AND SCIENCE,CBE-50.

1 Computer Science, SNMV College of Arts And Science,Cbe-50, (India)

2 Computer Science, SNMV College of Arts And Science,Cbe-50,(India)

ABSTRACT

*The abstract entitled "**Clinical Health Risk Prediction of Chronic Diseases from Sequential Pattern Mining** ". In the medical field, **chronic diseases** have been among the major concern which need prior assessment and detection. In this paper, a **novel system** for early evaluation on chronic diseases by mining successive risk and disease stage detection from analytic clinical records using effective sequential rules mining and classification methods.*

*The proposed system collects patient's medical history as input and predicts the risk of chronic disease with the condition score, which may help to the clinical team to select optimal solution. The proposed system performs effective **pre-processing, sequential risk detection**, classification with handling the class imbalance problem, and finally performs the post analysis and decision making process for the patient electronic health records.*

*The proposed system uses **UCI repository data** and performs risk prediction on health and liver chronic disease datasets. The system performs resampling process to effectively overcome the class imbalance problem.*

The results and analysis are performed using Weka and C#.net tool which provides the comparison analysis with accuracy, prediction time delay, and accuracy in decision selection etc.

I.INTRODUCTION

Data and information have become major assets for most businesses. Knowledge discovery in medical databases is a well-defined process and data mining an essential step. Databases are collections of data with a specific well defined structure and purpose. The programs to develop and manipulate these data are called DBMS. Knowledge discovery in databases is the overall process that is involved in unearthing knowledge from data.

International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 12, December 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

Data mining is concerned with the process of computationally extracting hidden knowledge structures represented in models and patterns from large data repositories. In paper [1] author defines KDD as a non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data. According to this definition, data are any facts, numbers or text that can be processed by a computer. The term pattern indicates models and regularity which can be observed within the data. The patterns, associations or relationships among all this data can provide information and it can be converted into knowledge about historical patterns and future trends. There are other steps such as data preprocessing, data selection, data cleaning and data visualization which also part of the KDD process.

Data mining is an interdisciplinary field of study in databases, machine learning and visualization'. It helps to identify the patterns of successful medical therapies for different illnesses and also it aims to find useful information from large collections of data .

Data mining is the core of KDD which is used to extract interesting patterns from data that are easy to perceive, interpret and manipulate. It is the science of finding patterns in huge reserves of data, in order to generate useful information. The KDD process comprises of few steps leading from raw data collections to form new knowledge.

As shown in Figure 1.1, described the knowledge discovery process as consisting of an iterative sequence of data cleaning, data integration, data selection, data mining pattern recognition and knowledge presentation. Data mining is the search for the relationships and global patterns that exist in large databases hidden among large amounts of data. A target data set must be assembled before data mining algorithms can be used. A common source for data is a data mart or data warehouse and pre-processing is essential to analyze these multivariate data sets. The final step of knowledge discovery from data is to verify that the patterns produced by the data mining algorithms occur in the wider data set.

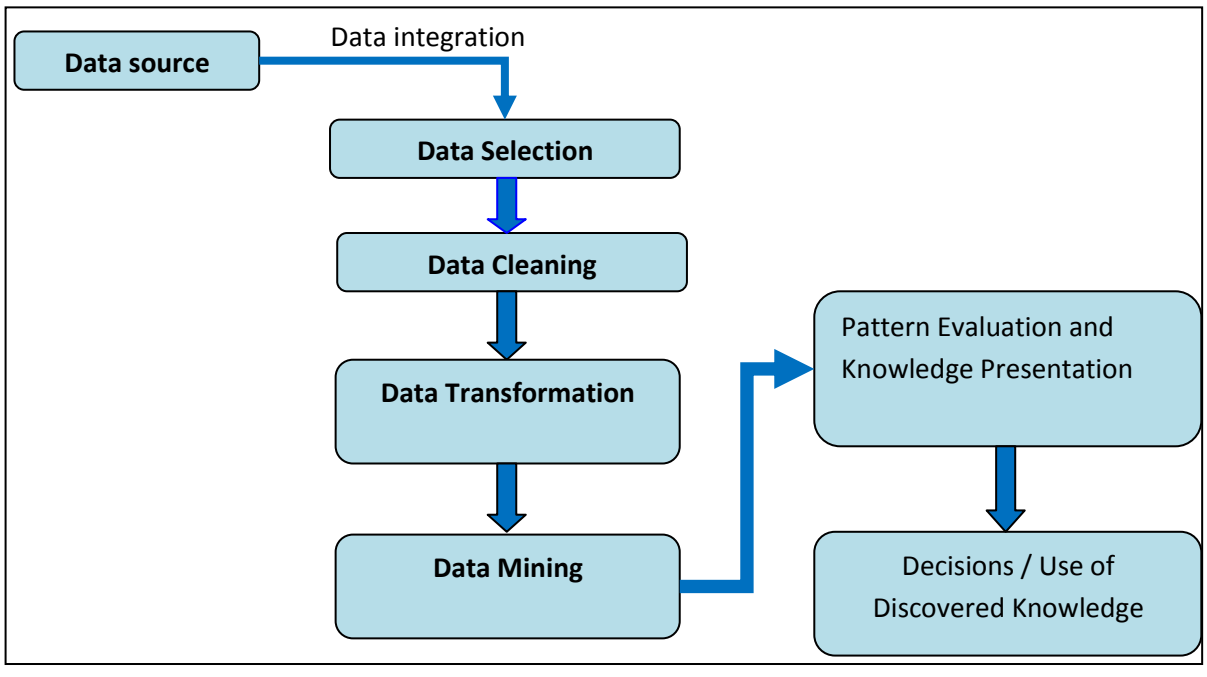

Figure 1.1 Knowledge discovery process

The discovered knowledge may contain rules that describe the properties of the data, patterns that occur frequently and objects that are found to be in clusters in the database etc., the motivation for handling data and performing computation is the discovery of knowledge. The KDD process employs data mining methods to identify patterns at some measure of interestingness and it is the process of turning the low-level data into high-level knowledge.

II. HEADINGS

1. Introduction

1.1. Data And Information Mining
1.2. Data Mining Tasks
1.3. Data Mining In Health Care
1.4. Data Mining Classification Techniques
1.5. Biological Link Between Diabetes And Heart Disease
1.6. Problem Definition
1.7. Objectives Of The Research
1.8. Organization Of The Thesis

2. Background Study

2.1. Introduction To The Risk Prediction
2.2. Data Mining In Healthcare Management
2.3. Existing System

3. Proposed System And Research Methodology

3.1. Proposed System And Its Contributions
3.2. Research Methodology

4. Implementation And Results

4.1. Dataset
4.2. Experimental Results
4.3. Performance Evaluation
4.4. Comparison Process

III. INDENTATIONS AND EQUATIONS

IV. FIGURES AND TABLE

Figure2.1 steps involved in the health risk prediction process

Figure 3.1 Proposed system architecture

Figure 3.2 sequential data samples from the diagnostic records

Figure 3.3 descriptive analysis results for the disease dataset

Table 2.1 data mining algorithms are proposed for heart disease

Table 2.2 data mining algorithms are proposed for Diabetes disease

V. CONCLUSION

The study proposed a new chronic disease prediction and sequential data mining scheme with different types of data's. The system studied the main two problems in the literature, which are sequential pattern mining time and classification delay. The study overcomes the above two problem by applying the effective enhanced LAPIN approach with intensive fuzzy algorithm. The LAPIN represents with the effective splitting criteria which has been verified by the INF algorithm. The system effectively identifies the disease from the sequential codes and its risk patterns, the sub type which is referred as the percentage of class such as normal and disease.

In summary, in this work, the main emphasis is placed on discovering the sequential risk patterns, which are well interpretable to predict the health risk of a new unknown instance.

The experimental results are evaluated using the C#.net and Weka. The experimental result shows that integrated extended LAPIN with INF algorithm shows better quality assessment compared to traditional fuzzy techniques. From the experimental results, the execution time calculated for classification object is almost reduced than the existing system.

VI. ACKNOWLEDGEMENTS

REFERENCES

[1] I. Kononenko, "Machine learning for medical diagnosis: History, state of the art and perspective," *Artif. Intell. Med.*, vol. 23, no. 1, pp. 89–109, 2001.

[2] G. D. Magoulas and A. Prentza, "Machine learning in medical applications," *Mach. Learning Appl. (Lecture Notes Comput. Sci.)*, Berlin/Heidelberg, Germany: Springer, vol. 2049, pp. 300–307, 2001.

[3] L. Breiman, "Bagging predictors," *Mach. Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[4] Gordan.V.Kass(1980). An exploratory Technique for inverstigation large quantities of categorical dataApplied Statics, vol 29, No .2, pp. 119-127.

[5] Leo Breiman, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone (1984). Classification and Regression Trees. Wadsworth International Group, Belmont, California.

[6] Quinlan J. R. (1986). Induction of decision trees. Machine Learning, Vol.1-1, pp. 81-106.

[7] Zhu Xiaoliang, Wang Jian YanHongcan and Wu Shangzhuo(2009) Research and application of the improved algorithm C4.5 on decision tree.

[8] Prof. Nilima Patil and Prof. Rekha Lathi(2012), Comparison of C5.0 & CART Classification algorithms using pruning technique.

[9] Baik, S. Bala, J. (2004), A Decision Tree Algorithm For Distributed Data Mining

[10]. V. Chauraisa and S. Pal, "Data Mining Approach to Detect Heart Diseases", International Journal of Advanced Computer Science and Information Technology (IJACSIT),Vol. 2, No. 4,2013, pp 56-66.

[11]. C. Aflori, M. Craus, "Grid implementation of the Apriori algorithm Advances in Engineering Software", Volume 38, Issue 5, May 2007, pp. 295-300.

[12]. Srinivas, K., "Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques", IEEE Transaction on Computer Science and Education (ICCSE), p(1344 - 1349), 2010.

[13]. Shanta kumar, B.Patil,Y.S.Kumaraswamy, "Predictive data mining for medical diagnosis of heart disease prediction" IJCSE Vol .17, 2011

[14] A. J.T. Lee, Y.H. Liu, H.Mu Tsai, H.-Hui Lin, H-W. Wu, "Mining frequent patterns in image databases with 9DSPA representation",Journal of Systems and Software, Volume 82, Issue 4, April 2009, pp.603-618.

[15] S. Bernard, L. Heutte, and S. Adam, "On the selection of decision trees in random forests," in *Proc. IEEE-ENNS Int. Joint Conf. Neural Netw.*, 2009, pp. 302–307.