Abnormal Drug Reaction Finding Using Temporal Nodes Bayesian Model

V. Livincy[1] M. N. Karuppusamy[2]
[1]P.G Scholar [2]Assistant Professor
[1,2]Department of Computer Science & Engineering
[1,2]Sri Subramanya College of Engineering and Technology, Palani Dindigul, India

Abstract—Drugs are frequently prescribed to patient's with the aim of improving each patient's medical state. But unfortunately most of the drugs produce undesirable side effects. The ADR method is not possible to investigate all the different drug combinations. The proposed system analyze the best drug and cross drug reaction and its symptoms based on association and temporal model. The proposed technique applies a prediction method with the existing temporal nodes Bayesian Model (TNBM) to data extracted from the patient and drug database in order to explore the probabilistic and best relationships between drug resistance mutations. Improving the classification and prediction accuracy, mutation and cross over functionalities has been proposed. The data mining algorithm CADR was developed to analyze the best and worst drug and drug pairs based on their casual and irregular reactions over a real electronic patient's database.

Keywords: Chronological Adverse Drug Reaction (CADR), DRESS, Temporal node Bayesian model, Association Rule Mining, P_GA

I. INTRODUCTION

Drugs are prescribed to patient's for improving each patient's medical state. But it may produce undesirable side effects. The side effects are either casually or infrequently occurs in the human body. Investigation have shown that the rate of unwanted side effects has been increasing annually and also no efficient way to identifying all the side effects of the drug. ADR is used to identify how the drug will react in the human body. When the side effect is detrimental to the patient's quality of life is reffered to as an ADE. But the ADR method is not possible to investigate all the different drug combinations. The rare ADR may unnoticed by medical practitioners and may cause morbidity or mortality in patient's. This could have been prevented by CADR method. This method can also help to detect many currently unknown ADR's.

Finding causal associations between two events or sets of events with relatively low frequency is very useful for various real-world applications. Negative side effects caused by prescribed medication currently present a huge burden for the healthcare service in terms of causing both patient morbidity and mortality and costing large sums of money. Investigations have shown that the rate of unwanted side effects has been increasing annually. Possible reasons for this are an increase in the number of annual prescriptions due to an aging population or an increase in polypharmacy, when numerous drugs are prescribed at the same time. Although it is common for a patient to develop side effects due to prescribed medication there is currently no efficient means of identifying all the side effects of a drug.

A. THIN Database:

The THIN database contains complete medical and prescription records for registered patients at participating general practices within the U.K. The medical information is recorded into the THIN database by Read Codes that correspond to illnesses, so each Read Code is paired with the illness description. Each Read Code is five elements long and the Read Codes have a tree structure.

Each drug is recorded into the THIN database by a drug code that is paired to the generic name. The drug code consists of nine numbers and does not have a structure we use but the drug code does specify the way the drug is ingested and the dosage. Each entry also includes the date that a Read Code or drug code is recorded but does not contain the time.

B. ADR:

When the side effect is detrimental to the patient's quality of life, it is often referred to as an adverse drug event (ADE) and when the drug causing the ADE is known, it is termed an adverse drug reaction (ADR).

ADRs can be discovered during the experimental Stages of a drug's development, but the occurrence of an ADR can depend on a magnitude of factors and it is impossible to investigate all the possible situations that may occur when the drug is taken. For example, testing for ADRs that result from polypharmacy would require clinical trials with millions of people to be able to investigate all the different drug combinations and this is not possible.

Current methods to discover rare ADRs often involve using a spontaneous reporting system (SRS) database that contains a collection of voluntary suspected drug and ADR reports.

The algorithms that signal ADRs by mining SRS Databases calculate a measure of how disproportionally more often the medical event is reported with a specific drug of interest compared with any drug.

II. RELATED WORK

Drugs and their associated CADRs have causal relationships. In this section, this examines how to search for potential CADR signal pairs from an electronic patient database using the above exclusive causal-leverage measure. This assumes that patient data are stored in relational tables in a database and can be retrieved using database language like structured.

A. Dataset Processing:

Data extractions are performed to feed a common repository. The extracted data have to fit the data model defined above. An important point is that no data have to be specifically recorded for the project as only routinely

All rights reserved by www.ijsrd.com

collected data are used. Those data include Medical and administrative information, e.g. age, gender, admission date, and medical department. This uploads patient and drug details. The patient details such as patient name, age, gender and symptoms. The drug details such as drug number, type of the drug and description about the drugs.

B. Association Finding:

The association finding module helps to track all the support and confidence of the drugs. This initially contains the following process. They are Scan the database, Candidate generation, Support and confidence calculation.

C. CADR Model:

The ADR model was preliminarily validated in the previous study by using it to calculate the extent of causality between symptoms and some of its adverse effects. This used the real patients to create simulated patient cases, all of which containing drug-symptom pairs of interest with various degrees of causality. The model's validity was then established by comparing the decisions made by the model and those by two independent experienced physicians for the set of simulated patients.

D. Pair Generation and Evaluation:

This is the process for pair generation and evaluation. In this algorithm drug-symptom pairs that can be easily generated. The pairs are drug-drug pairs, symptom-symptom pairs or combinations of multiple drugs and symptoms. Thus, this algorithm generates a much fewer number of candidate rules, which implies much less complexity.

E. Causal-Leverage:

This algorithm shows how to compute the causal-leverage value of a general pair between event X and Y. Both X and Y could be either drug event or symptom event. First, the drug or symptom hash table is searched in order to get the support count for event Y. Then, for each PID that supports the pair, a process called cue abstraction is used to extract a set cue values V from the related patient case. Specifically, a list of drug start dates and a list of symptom dates are retrieved from the Patient Drug Table and the Patient Symptom Table, respectively. Finally rank all the pairs in a decreasing order according to their exclusive causal-leverage values after all these values are computed.

F. Subset Finding and Result Analysis:

The final module has implemented to analyze the subset and coordinal relationships of drug pairs. The system performs the detailed analysis based on the temporal analysis of the drugs. Finally the system performs the analysis and results of the proposed system. This proposed a data mining algorithm to mine ADR signal pairs from electronic patient database based on the new measure. The algorithm's computational complexity is analyzed.

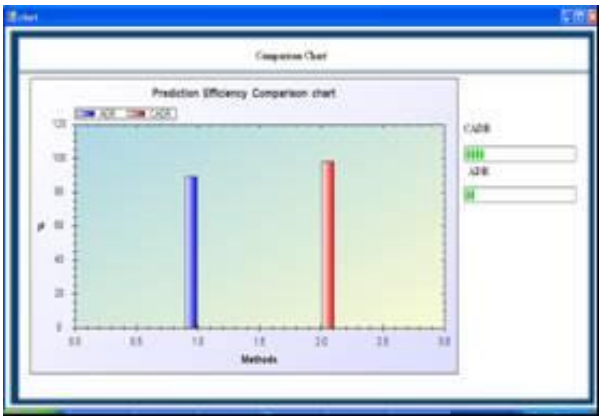

Fig. 1: Efficiency Comparision Chart

III. OBJECTIVES OF THE EXISTING MECHANISM

There are several approaches implemented in existing to detect the adverse of drug reactions. Several top down and bottom up approaches are proposed to detect of rare association rules based on traditional interestingness measures like support and confidence. One pitfall of these measures is that they simply find the statistical correlation between two points. They do not indicate any temporal relationship between X and Y. In addition, they are not able to capture the causal relationships between two event sets. Existing system may be challenged to express their knowledge in the form of probability distributions. But it failed to compare two different mixture treatments along with the temporal occurrence of drugs and drug challenging reactions and changes in order to predict the most effective treatment. Verification of each drug reaction based on various conditions is very difficult. The existing system provides only an approximate solution for treatment selection, so that is not an accurate system.

Since a patient database normally only contains a subset of all drugs on the market and a subset of all symptoms, it is necessary to search the Patient Drug Table and the Patient Symptom Table to get the drugs and symptoms covered by the database. Formula for the Traditional interestingness measures like support and confidence are given below

Support = [(no of dq in data set D/total no of dataset D)*100]

Confidence = [(AUF)/T(A)]*100

Since a patient database normally only contains a subset of all drugs on the market and a subset of all symptoms, it is necessary to search the Patient Drug Table and the Patient Symptom Table to get the drugs and symptoms covered by the database. Formula for the Traditional interestingness measures like support and confidence are given below

Support = [(no of dq in data set D/total no of dataset D)*100]

Confidence = [(AUF)/T(A)]*100

Systems design is the process of defining the architecture, components, modules, interfaces, and data for a system to satisfy specified requirements. Systems design could be seen as the application of systems theory to product development. The system architecture for best treatment detection is given below.

All rights reserved by *www.ijsrd.com*

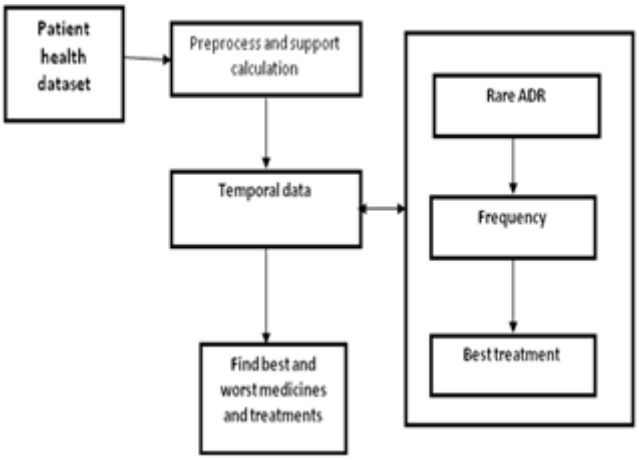

Fig. 2: Best Treatment Detection

IV. OVERVIEW OF THE PROPOSED MECHANISM

The system proposes an innovative data mining framework and applies it to mine potential causal associations in electronic patient data sets where the drug-related events of interest occur infrequently and temporally. The proposed system analyzed the best drug and cross drug reactions and its symptoms based on associations and temporal model. Aims to propose a data mining algorithm to mine CADR (chronological adverse drug reaction) signal pairs from electronic patient database based on the new measure.

The system performs the association rules for support calculation and performs the temporal mutations. The proposed technique applies a prediction method with the existing temporal nodes Bayesian Model (TNBM) model to data extracted from the patient and drug database in order to explore the probabilistic and best relationships between drug resistance mutations. Analyzing those rules by considering several temporally divided dataset with genetic approach is proposed. Improving the classification and prediction accuracy, mutation and cross over functionalities has been proposed. The proposed system performs P_GA, which is a prediction scheme with the slotted training dataset changing values.

Mining these associations is very difficult especially when events of interest occurs infrequently. This has developed a new interestingness measure which finds the exclusive causal-association and infrequent association based on an experience-based Genetic decision model. A data mining algorithm was developed to analyze the best and worst drug and drug pairs based on their casual and irregular reactions over a real electronic patient database for potential drug finding. The proposed system utilizes the synthetic dataset for the experiment.

V. P_GA ALGORITHM

Genetic Algorithm is a search heuristic that imitate the process of natural evolution. Genetic algorithms (GAs) are inspired in Darwin's theory of evaluation by natural selection and are powerful tool in difficult search and optimization problems. Genetic algorithm model the solution of the problem into data structure called chromosome or genotype which represents the possible solutions called individuals or phenotypes. A sequence of genetic operators is applied to these chromosomes in order to achieve a high optimization of the problem. Two components play a significant role in the GA method: the problem codification and the evaluation function. The problem codification is the mapping that is made between the chromosomes and the individuals. The evaluation function takes one individual and calculates its fitness. Usually, the fitness is a performance measure of the individual as a solution to the problem.

Genetic algorithms employ metaphor from biology and genetics to iteratively evolve a population of initial individuals to a population of high quality individuals, where each individual represents a solution of the problem to be solved and is compose of a fixed number of genes. The number of possible values of each gene is called the cardinality of the gene. Each individual is called as chromosome. The set of chromosomes forms population. Functioning of genetic algorithm starts with randomly generated population of individuals. Thru various generations these population evolved and individuals' quality gets improved.

In every generation, three basic operators of genetic algorithm i.e. selection, crossover and mutation are applied to each individual. Crossover means exchanging the genes between two chromosomes while mutation means random changing of a value of a randomly chosen gene of a chromosome. These individuals are representation of the problem required to be solved. Different positions of each individual can be encoded as bits, categorical and numerical Here, the numbers of best-fit individuals are selected. For this user defined fitness function is used. Fitness function is used to measure quality of each chromosome. Remaining individuals are paired and thru process of crossover, new offspring is produced by partially exchanging their genes. When genetic algorithm is used for problem solving, three factors will have impact on the effectiveness of the algorithm, they are,

The selection of fitness function, The representation of individuals and The values of the genetic parameters.

Genetic algorithm is used for evolving new population for outlier detection. Using these rules normal network traffic or audit data differentiated from abnormal traffic/data. Rules in the rule set of genetic algorithm are of type if-then. Following is general syntax for rule in genetic algorithm,

if { condition } then { act }

Condition refers to the data to be verified and rule in rule set while act is the action to be performed if condition is true. A condition can check for categorical attributes and values of new medical data's. While act refers to the action to be performed when condition is true like sending alert message, creating log messages etc. Benefits of using genetic algorithm for outlier detection are,

1) Genetic algorithms are intrinsically parallel. Because of multiple offspring, they can explore the solution space in multiple directions at once.
2) Parallelism allows genetic algorithm to implicitly evaluate many schemas at once. This make them well suited to solving problems where space of potential solution is truly huge.
3) Genetic algorithm based systems can be re-trained easily. This improves its possibility to add new rules and evolve outlier detection system.

All rights reserved by *www.ijsrd.com*

A. Genetic Algorithm for Fitness Function:

1) A constant size population of individuals
2) Each individual represents a point in the search space for a given problem through a suitable coding. A fitness value is assigned to each individual in the population
3) Individuals are ranked and selected according to their fitness in such a way that more fit individuals are more likely to enter the relevancy group
4) Genetic operators such as crossover and mutation are applied to pairs of individuals or single individual in order to produce new individuals.

The chapter represents the best fit calculation of attributes based on the genetic approach the calculation passes the mild and extreme outlier ranges for the best fir as well as worst fit calculation

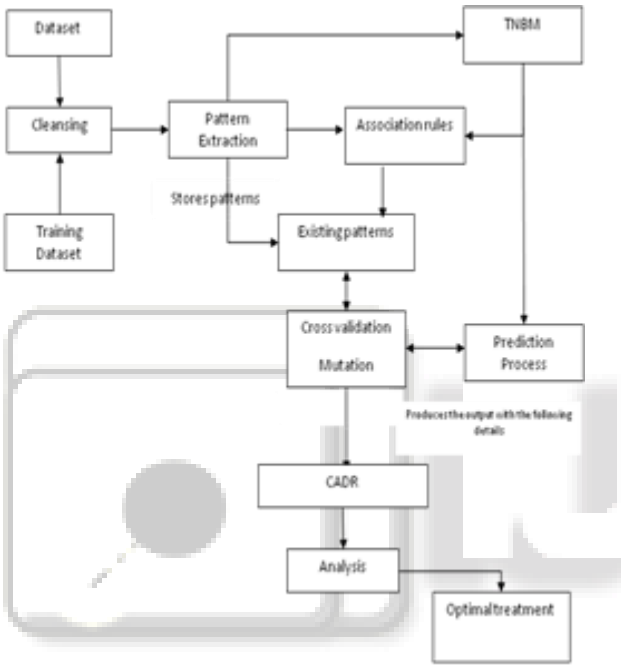

Fig. 3: Expected Outcomes

VI. CONCLUSION

This proposed work was implemented using C#.net. The performance of this proposed work CADR using association rule mining and Genetic with TNBM Scheme was compared with the existing ADR. The main contribution of the paper is to use a temporal Bayesian method and chronological adverse drug reaction approach to understand different drug reaction changes. The proposed system has implemented for best and worst drug findings some important known correlated mutations were discovered, as well as other temporal relations.

The system also compared the TNBN approach with genetic models such as Bayesian networks and association rules. Finally the system compared with existing models. The proposed system obtained important information that the accuracy level.

REFERENCES

[1] Huidong Jin, Jie Chen, Chris Kelman, Hongxing He,Damien McAullay, and Christine M. O'Keefe, "Mining Unexpected Associations for Signalling Potential Adverse Drug Reactions from Administrative Health Databases".
[2] Andreea Farcas and Marius Bojita, "Adverse Drug Reactions in Clinical Practice: a Causality Assessment of a Case of Drug-Induced Pancreatitis".
[3] Yihui Liu and Uwe Aickelin, "Detect adverse drug reactions for the drug Pravastatin".
[4] Roseleen Vino.I and Kerana Hanirex.D "Sharing ADRs for Immediate Treatment".
[5] R. Sindhulakshmi and, D. SaravanaPriya, "A Data Mining approach for efficient event analysis to find Adverse Drug Reactions".
[6] S. Prakash and S. Kanjanadevi,"Post Market Drug Analysis using Irregular Pattern Mining Scheme".
[7] S.Sandhya Lakshmi and S.Sheba Roshni,"Mining Adverse Drug Reaction For Infrequent Causal Association".
[8] G. Shepherd, P. Mohorn, K. Yacoub, and D. W. May, "Adverse drug reaction deaths reported in united states vital statistics, 1999–2006," Ann. Pharmacother., vol. 46, no. 2, pp. 169–175, 2012.
[9] T. M. Betteridge, C. M. Frampton, and D. L. Jardine, "Polypharmacy- We make it worse! a cross-sectional study from an acute admissions unit," Internal Med. J., vol. 42, no. 2, pp. 208–211, 2012.
[10] M. Pirmohamed, S. James, S. Meakin, C. Green, A. K. Scott, T. J.Walley, K. Farrar, B. K. Park, and A. M. Breckenridge, "Adverse drug reactions as cause of admission to hospital: Prospective analysis of 18 820 patients," Brit. Med. J., vol. 329, no. 7456, pp. 15–19, 2004.
[11] F. R. Varallo, M. F. R. Lima, J. C. F. Galdur´oz, and P. C. Mastroianni, "Adverse drug reaction as cause of hospital admission of elderly people: A pilot study," Latin Amer. J. Pharmacy, vol. 30, no. 2, pp. 347–353,2011.
[12] J. A. Berlin, S. C. Glasser, and S. S. Ellenberg, "Adverse event detection in drug development: Recommendations and obligations beyond phase 3," Amer. J. Public Health, vol. 98, no. 8, pp. 1366–1371, 2008.

All rights reserved by *www.ijsrd.com*