RESEARCH ARTICLE OPEN ACCESS

DEEP LEARNING PREDICTION OF ADVERSE DRUG REACTION ANALYSIS USING ARTIFICIAL NEURAL NETWORK MODEL

Mrs. K.E. Eswari, M.C.A., M.Phil., M.E., SET.,[1], R. Sarathkumar[2],

[1]Associate Professor, Department of Computer Applications, Nandha Engineering College (Autonomous), Erode, Tamilnadu, India.

[2]Final MCA, Department of Computer Applications, Nandha Engineering College (Autonomous), Erode, Tamilnadu, India.

Email: [1]eswarisaravanan2001@gmail.com, [2]rajasarath28@gmail.com

----------**********----------

Abstract.

In medical domain, Adverse Drug Reaction (ADR) analysis is a crucial process for doctors and medical scientists.Adverse drug reaction measures the injury occurred due to usage of a drug. The growing concern to the ADRs hasstimulated the progress of statistical, data mining methods to find the Adverse Drug Reactions. This project proposed ahybrid model of data mining and machine learning to classify different Adverse Reactions and foretell the outcomeintensity. It used the Proportionality Reporting Ratio (PRR) along with Chi-Square test equations to find out the different relationships between drug and symptoms called the drug-ADR association. In addition, support vector machine method is applied to classify the data set records into either normal or adverse drug. Moreover, our project aims in finding the percent of adversity of the drug reaction. Based on the number of occurrences, where a specific drug P, causes a specific Adverse Reaction, R, the various terms are measured for PRR and Chi Square. In addition KNN andSVM classification is made on drug records to classify them based on data set columns. Neural network based classification is the proposed system to classify drug reaction based on data set columns.

Keywords: Adverse Drug Reaction, Proportionality Reporting Ratio, Chi Square, Neural Network.

I. INTRODUCTION

This project used the Proportionality Reporting Ratio (PRR) along with the precision point estimator test called the Chi-Square test to find out the different relationships between drug and symptoms called the drug-ADR association. Thisoutput is used as an input to machine learning algorithms such as Random Forest and Support Vector Machine (SVM)to predict the intensity of the outcomes of ADR. The aforementioned challenges motivated a series of works that apply data mining and machine learning approachesto come up with various solutions using different datasets to suit according to the researcher's needs. Google recentlyworked on a Twitter dataset to detect ADEs from posts on Twitter using a merged form of Artificial Neural Network(ANN).

ANN uses a binary classifier to represent outcome result. Another researcher proposed the Predictive PharmacosafetyNetworks (PPNs) for detecting unknown ADEs. The existing drug safety information is used from a well-known data setof drug safety in 2005 to train a logistic regression model to detect unknown ADEs. In 2012, a research used the 'THIN'database to create a model using the feature matrix and feature selection to identify ADRs for a specific drug called"Pravastatin".

In the world of medical science and drugs, Adverse Drug Reaction (ADR) has always been an important field ofresearch. Adverse drug reaction means the injury from the use of a drug. These injuries can extend from minor injurieslike skin rash to major life-threatening reactions.

Confusion occurs mostly between ADR and Side Effect where ADR is the reaction caused by the drug used at normaldoses for particular symptoms. Point to be noted: a wrongful overdose of drugs is not considered as an ADR case.

Every Year more than 200 thousand deaths are reported because of ADRs. Though ADR can be identified very easilyafter the occurrence, predicting ADR has always been a huge challenge for researchers.Worldwide, around 4.9% of hospital admissions are the result of ADRs and this number is as high as 41.3% in someareas. In Sweden, ADRs are the seventh most common cause of death. Even though drugs are thoroughly testedclinically before they are

released on the market, many unknown side effects are discovered after they have been usedover time by various patients. Moreover, as ADR varies from person to person, predicting ADR can sometimes be ashard as impossible even for doctors.

Nowadays almost 73% of the people around the world take different medications. Among these, almost 29% of thesemedicines have different kinds of adverse drug reactions. FDA statistics show that almost 7000 of these ADRs havecaused death in recent years.

This project tried to predict the ADR and its intensity so that necessary precautions can be taken before prescribing anykinds of medications. In recent years, a huge amount of Electronic Medical Records (EMR) is available on severalplatforms. As a result, various methods of data science can be implemented for the detection of ADR.

Various methods of machine learning and data mining have already been implemented to predict the possible ADRfrom a specific drug. However, this project proposed a hybrid model of machine learning and data mining to predict theADR and its severity. We contributed mostly on the machine learning part where we took the detected ADR from thealready existing, statistical data mining techniques and usedit to find the severity of the reaction on various uniquepatients.

II. LITERATURE REVIEW

Adverse Medicine events beget substantial morbidity and mortality, yet they remain uncredited and misknew. Thelanguage to describe crimes and patient detriment associated with specifics causes important confusion. Thiscomposition uses the case study of a case with multiple adverse medicine events to clarify crucial terms, similar asadverse event, adverse medicine response, adverse medicine event, drug error, and side effect. The case discussionillustrates clinical approaches to assaying the unproductive connection between a suspect medicine and an adverseevent. Exemplifications and explanation for meaningful attestation of adverse medicine events are handed, along withan figure of the types of events that should be reported to nonsupervisory agencies.

Since the early 1990s, adverse medicine events have entered significant attention from experimenters in quality andpatient safety (11). Nationally honored quality experts have linked adverse medicine events as a top safety precedence(12) because these events are the most common type of iatrogenic injury (13). Studies have indicated that adversemedicine events do nearly daily in medium-sized hospitals and inpatient panels (14 – 16).

Still, despite the high morbidity and mortality, croakers frequently don't fete or meetly treat cases of medicine-relateddetriment (17, 18). The authors believed that shy recognition and treatment of medicine- related detriment are, in part, aresult of what has been called a Palace of Babel of language (11). Terms firstly developed in the narrow environment ofmedicine goods in a clinical and nonsupervisory setting are now being applied in the broader environment of qualityenhancement in health care delivery systems (19).

As might be anticipated, the expanding part of these terms has been coupled with their use in antithetical ways, indeedwithin the same discipline (20). In this paper, they used the case of an factual case as a frame to explain therecognition, treatment, attestation, and reporting of medicine- related detriment.

ADVERSE EVENTS VERSUS ADVERSE Medicine ResponsesMr.J. was a 70- time-old man with nephrotic pattern, pneumoconiosis, and a history of gout and myocardial infarction. Hepresented to the sanitarium with adding bilateral leg edema and pain, for which he'd been taking untoward ibuprofen,400 mg three times a day for 3 days and once a day for the antedating 3 weeks.His other inpatient specifics were simvastatin, 40 mg at bedtime; aspirin, 81 mg formerly daily; and metoprolol, 50 mgdoubly daily. In the exigency department, his serum creatinine position was 680 spook/ L (7.7 mg/ dL), much advancedthan the birth of 290 spook/ L (3.3 mg/ dL) 11 months before. He was admitted to the sanitarium. The case endured anadverse event while using ibuprofen.

Is this event a side effect, an adverse medicine response, a drug error, or an exacerbation of his underpinning renaland cardiac complaint? Terms that originally arose from the field of pharmacovigilance, similar as adverse event andadverse medicine response, can help croakers relate the edema and renal failure to ibuprofen. Pharmacovigilance isthe study of medicine-related injuries for the purpose of making warning or pullout recommendations forpharmaceutical products. The International Conference on Harmonisation of Technical Conditions for Registration ofMedicinals for Human Use, of which theU.S. Food and Drug Administration (FDA) and the World Health Organizationare members, defines an adverse event as “ any untoward medical circumstance that may present during treatmentwith a pharmaceutical product but which doesn't inescapably have a unproductive relationship with this treatment”.

In this paper (2) the authors studied the problem of detecting rulings describing adverse medicine responses (ADRs)and frame the problem as double bracket. They delved different neural network (NN) infrastructures for ADR bracket. Inparticular, we propose two new neural network models, Artificial Intermittent Neural Network (CRNN) byconcatenating Artificial neural networks with intermittent neural networks, and Artificial Neural Network withAttention ANNA by adding also the attention weights into ANN. They estimated colorful NN infrastructures on a Twitter dataset containing informal language and an Adverse Drug Goods (ADE) dataset constructed by slice from MEDLINE case reports.

Experimental results show that all the NNinfrastructures outperform the traditional maximum entropy classifiers trained from n-grams with different weightingstrategies vastly on both datasets. On the Twitter dataset, all the NN infrastructures perform also. However, in the ADEdataset, neural network does better than other more complex ANN variants. Nonetheless, ANNA allows the

visualization of attention weights of words when making bracket opinions and hence is more applicable for the birth ofword subsequences describing ADRs. Adverse Medicine Responses (ADRs) are potentially veritably dangerous to cases and are amongst the top causes ofmorbidity and mortality. Numerous ADRs are hard to discover as they be to certain groups of people in certainconditions and they may take a long time to expose. Healthcare providers conduct clinical trials to discover ADRsbefore dealing the products but typically are limited in figures.

Therefore, post-market medicine safety monitoring is needed to help discover ADRs after the medicines are vended on the request. In the United States, Robotic Reporting Systems (SRSs) is the sanctioned channel supported by the Foodand Drug Administration. Still these system are generally under- reported and numerous ADRs aren't recorded in thesystems.

Lately unshaped data similar as medical reports or social network data have been used to descry content that containsADRs. Case reports published in the scientific biomedical literature are abundant and generated fleetly. Social networks are another source of spare data with unshaped format. If the individual tweet or status in Facebook contains ADRs, it may not be clinically useful, and so a large volume of these data can expose unknown or serious consequences. Common approaches to descry content with ADRs used Support Vector Machines (SVMs), Random Forest, Maximum Entropy classifiers with heavily hand- finagled features. These features typically include n-grams with different weighting schemes.

When used with unigrams, these approaches suffer from the fact that their models don't take in account the commerce between terms and their orders. This problem can incompletely be answered by using bi-grams or trigrams. Still this leads to the number of features exploding, and the models are therefore fluently overfitted. Meanwhile neural networks with pre-trained word representations have had some successes in other textbook bracket tasks.

Word representations that are generally pre-trained with unlabelled data are matrices that can be used to project words into a thick low-dimensional space (generally from 50 to 300 confines). These neural networks of ten contain Artificial pollutants or intermittent connections that cipher weighted totalities of words and their surrounds. The authors in this paper, trained a) word embeddings and b) use them as parameters to different neural network infrastructures for classifying documents to whether they contain ADR content or not.

They concluded that this paper has explored different neural network (NN) infrastructures for ADR bracket. In particular, they have proposed two new neural network models, a) Artificial Intermittent Neural Network (CRNN) and b) Artificial Neural Network with Attention (ANNA). Experimental results verify that all the NN infrastructures outperform traditional Maximum Entropy classifiers trained from "n-grams" with varying weighting strategies vastly on both Twitter and ADE datasets.

Among NN infrastructures, no significant differences were observed on the Twitter dataset. But ANN seems to do better when compared with other more complex ANN variants on ADE dataset. Nonetheless, ANNA allows the visualization ofattention weights of words when making bracket opinions and hence is more applicable for the birth of word subsequences describing ADRs.

The authors of this research [3] noted that adverse drug reactions (ADR) are a major public health concern. They suggesteda novel approach to detect ADRs using a feature matrix and feature selection in this work. The tests are conducted on themedication Pioglitazone.

When compared to other automated systems, major negative effects for the drug are discovered,and better

performance is attained. The observed ADRs were discovered using an automated method, and more researchis required.

Adverse drug reactions (ADR) are a major public health concern. ADRs are one of the most common reasons whymedications are withdrawn from the market. Thespontaneous reporting system (SRS) and prescription event monitoringare now the two main strategies for detecting ADRs (PEM).

In pharmacovigilance, a signal is defined as "any published information on a putative causative association between anadverse event and a medicine, the relationship being unknown or incompletely documentedpreviously," according to the World Health Organization (WHO).Many machine learning methods, such as Bayesianconfidence propagation neural network (BCPNN), decision support method, genetic algorithm, knowledge based approach, and others, are employed to detect ADRs in the spontaneous reporting system.

One shortcoming is the reporting method for submitting ADR reports, which suffers from significant underreporting and is unable to adequately evaluate the risk. Another problem is that ADRs are difficult to detect in databases with a small number of occurrences of each drug-eventconnection. We present a feature selection strategy for detecting ADRs in The Health Improvement Network (THIN)database in this research.

By connecting patients' prescriptions and accompanying medical occurrences, the first feature matrix is constructed, which reflects medical events for patients before and after taking medications. Then, using featureselection methods, significant characteristics are picked by comparing the feature matrix before and after patients takemedications.

Finally, based on comparable traits, substantial ADRs can be discovered from thousands of medicaloccurrences.The drug Pioglitazone is the subject of research. A good result is obtained. The first feature matrix is taken from the THINdatabase, which summarises the medical events that patients experience before or after taking medications, in order todetect drug ADRs.

Then, using the Student's t-test feature selection approach, significant features are selected from a feature matrix containing thousands of medical occurrences.The authors concluded that using a feature matrix and feature selection, they provided a unique method for successfullydetecting ADRs.

The THIN database is used to generate a feature matrix that defines medical occurrences before and afterindividuals take medications.

The Student's t-test feature selection approach is used to identify relevant features inthousands of medical events. Significant ADRs are identified, which connect to significant traits. The drug Pioglitazone isthe subject of research. In comparison to existing automated methods, the proposed method performs well

III. PROPOSED METHODOLOGY

In existing system, to find an association between the drug and the symptoms accountable for the prescribed drug, datamining association techniques such as Chi-Square and Proportionality Reporting Ratio (PRR) are used which givenumerical values for the model. If the values exceed a certain threshold, that specific symptom is identified as an ADRfor a certain drug. The general conditions for running the PRR are as mentioned below:

- Value A refers to the number of occurrences, where a specific drug P, causes a specific Adverse Reaction, R.
- Value B refers to the number of occurrences, where specific drug P causes any other Adverse Reactions but R.
- Value C refers to the number of occurrences, where the Adverse Reaction R is caused by any other drug but P.
- Value D refers to the number of occurrences, where any other Adverse Reactions but R is caused by any other drugs but P

PRR= (A / (A+B)) / (C / (C+D))

Chi Square X2 = (AD-BC)2+(A+B+C+D)/[(A+B) (B+C) (C+D) (A+D)]

The threshold value for Chi-Square and PRR:

• PRR > 2

• Chi-Square >4.

The drug reactions with values greater than the threshold are matched with already given symptoms by the

user.Ifthere is any cross match, then that is considered as a potential ADR. Drawbacks are:

- Decides only if the drug is fully adverse or not.
- Threshold value is not pre-decided.
- The existing system surely does not give precise results.
- Based on the PRR values, doctors need to decide themmanually for ADR or not.
- KNN and SVM classification is not made on drug records to classify them.

In proposed system, all the existing approaches are carried out. In addition, SVM classification is carried out in which80% of the data are given as training data. The 20% of the data are taken as test data. KNN and SVM classification isalso made to classify the drugs based on dataset columns.

The probability value attribute of the SVM model is set totrue and so the test data record yields a numeric value between zero and one which indicates the level of adversity ofthe drug for that record values. So, unlike existing system which gives either ADR or not, here the percentage ofadversity could be found out. Advantages are:

- Decides the level of drug adversity.
- Threshold value is not required.
- The proposed system gives precise results. Based on the SVM values, ADR level could be found out.
- KNN and SVM classification is made on drug records to classify them based on data set columns.

IV. FINDINGS

To achieve Adverse Drug Reaction analysis results, the dataset itself should be improved in so many ways.

A patient'sblood group, diabetes information, pregnancy condition (female patients) can be very important factors in determiningspecific ADRs or to predict the outcome of an ADR. This information is missing from our dataset.

So, if a new datasetcan be formed comprising of the valuable information and our above-mentioned concepts are applied to the dataset, amuch better result can be achieved.

Secondly, since ADRs are person specific mostly and can vary from patient topatient, a groundbreaking result can be achieved if the genetic diagnosis information is available for all the patients andcan be used in determining the ADRs.

Moreover, if information about the drugs is available on a molecular level, they can be aided with the person's geneticinformation to create a more stable system for the prediction of Adverse Drug Reaction and their intensity.

If the newsystem, applies KNN and SVM prediction model, then the prediction will be adoptable for future new drug records.

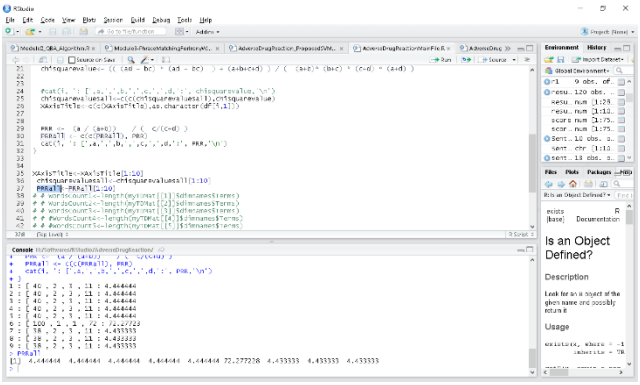

FIG 4.1 PROPORTIONALITY REPORTING RATIO VALUES

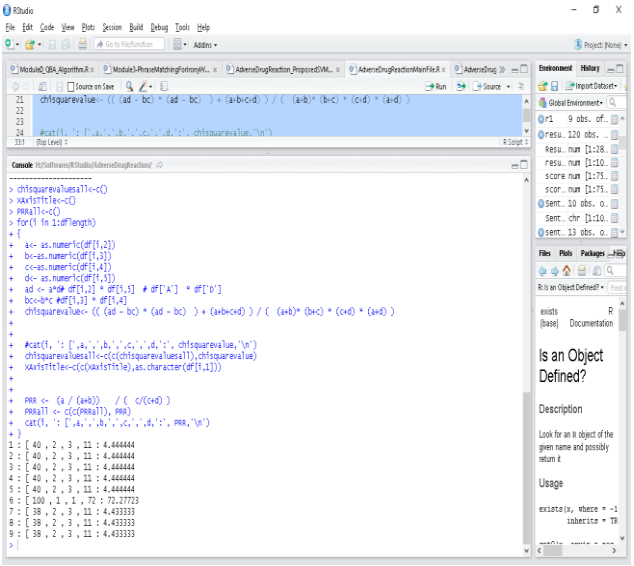

FIG 4.2 CHI SQUARE VALUES

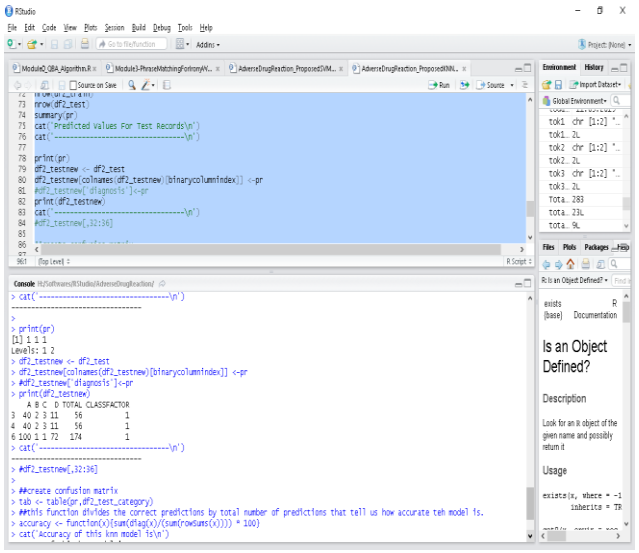

FIG 4.3 KNN CLASSIFICATION BASED PREDICTED VALUES

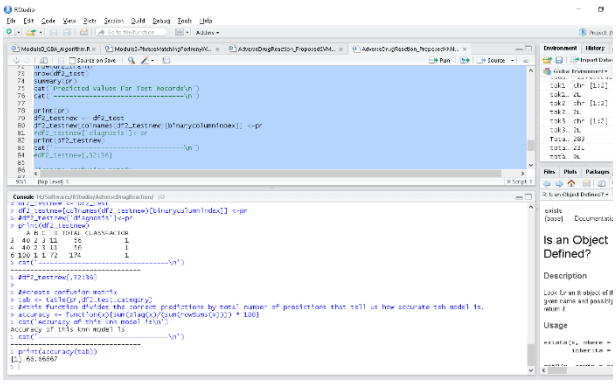

FIG 4.4. ACCURACY OF THE KNN MODEL

V. CONCLUSION

The project has explored various concepts of data mining and machine learning and attempted to come up with ahybrid model that will help doctors and pharmacists to perform a safe drug evaluation on a combination of drugs beforethey prescribe medicine to the patients.The Proportionality Reporting Ratio (PRR) and Chi-Square test are used as the data mining technique to help evaluate the correct combination of safe drugs to be prescribed to the patient and also, it went further ahead with the proposedmachine learning concepts to help doctors and pharmacist to be well aware of the outcome of an Adverse Event if itwere to occur from a combination of drugs. A drug is considered safe until an adverse reaction is reported for that drug.However, in the field of medicine, there is always scope for uncertainty since each drug can react differently to differentspecific patients. This system can be used as a complimentary tool with the doctor's knowledge and can help aid themin performing a safe drug diagnosis and prescribe the correct combination of medicine to the patient.In addition, this project classifies the drug data set records using KNN and neural network to predict the model forfuture test record data sets. It will be helpful in analyzing the drug in the current data set and so as to predict the futureupcoming dataset. In future, logistic regression can be applied to further classify the data.

REFERENCES

[1] J. Nebeker, P. Barach and M. Samore, "Clarifying Adverse Drug Events: A Clinician's Guide to Terminology, Documentation, and Reporting", Annals of Internal Medicine, vol. 140, no. 10, p. 795, 2004.

[2] T. Huynh, Y. He, A. Willis and S. R¨uger, "Adverse Drug Reaction ClassificationWith Deep Neural Networks", in Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, Osaka, Japan, 2016, pp. 877-887.

[3] Y. Liu and U. Aickelin, "Detect adverse drug reactions for drug Pioglitazone", 2012 IEEE 11th International Conference on Signal Processing, 2012.

[4] L. Duan, M. Khoshneshin, W. Street and M. Liu, "Adverse Drug Effect Detection", IEEE Journal of Biomedical and Health Informatics, vol. 17, no. 2, pp. 305-311, 2013.

[5] A. Tripathy, N. Joshi, H. Kale, M. Durando and L. Carvalho, "Detection of adverse drug events through data mining techniques", in 2015 International Conference on Technologies for Sustainable Development (ICTSD), Mumbai, India, 2015, pp. 01-06.

[6] Ross SD. Drug-related adverse events: a readers' guide to assessing literature reviewsandmeta-analyses.ArchInternMed.2001;161:1041-6.[PMID:11322836]

[7] Shojania KG, Duncan BW, McDonald KM, Wachter RM. Safe but sound: patient safety meets evidence-based medicine [Editorial]. JAMA. 2002;288:50813. [PMID: 12132985]

[8] Kohn LT, Corrigan JM, Donaldson MS. To Err Is Human: Building a Safer Health System. Washington, DC: National Academy Pr; 1999.

[9] Bates DW, Cullen DJ, Laird N, Petersen LA, Small SD, Servi D, et al. Incidence of adverse drug events and potential adverse drug events. Implications for prevention. ADE Prevention Study Group. JAMA. 1995;274:29-34. [PMID: 7791255]

[10] Gandhi TK, Weingart SN, Borus J, Seger AC, Peterson J, Burdick E, et al. Adverse drug events in ambulatory care. N Engl J Med. 2003;348:1556-64. [PMID: 12700376]

[11] Fattinger K, Roos M, Vergeres P, Holenstein C, Kind B, Masche U, et al. Epidemiology of drug exposure and

adverse drug reactions in two swiss departments of internal medicine. Br J Clin Pharmacol. 2000;49:158-67. [PMID: 10671911]

[12] Beatrice A. Golomb, John J. McGraw, Marcella A. Evans and Joel E. Dimsdale. Physician Response to Patient Reportsof Adverse Drug EffectsImplications For Patient-Targeted AdverseEffect Surveillance. Drug Safety 2007; 30 (8): 669-675. 0114-5916/07/0008-0669/$44.95/0.