

Effective Pharmacovigilance using Natural Language Processing and Neural Network

T.Suresh Balakrishnan¹, Dr. Latha Parthiban²,

¹Ph.D Research Scholar, Dept. of Computer Science & Engineering

St Peter's Institute of Higher Education and Research Avadi, Tamilnadu, India.

²Department of Computer Science, Pondicherry University CC, Puducherry, India.

Corresponding author Email : lathaparthiban@yahoo.com

Abstract

Pharmacovigilance plays a vital role in monitoring the Adverse Drug Reaction (ADR) caused due to single drug intake, combined dose and prolonged administration. ADR has led to increase in mortality rate by 1.8% throughout the world. Until now monitoring of ADR is a challenging task due to self-medication in countries like India and improper physician follow up. In this paper, a novel method is proposed to evaluate the ADR using online and offline reviews through natural language processing and neural network. From the result obtained, an effective way to improve the ADR monitoring is identified and the gap between the patients, physician and pharmacists is reduced.

Keywords: ADR, NLP, self-medication, online and offline reviews

I. INTRODUCTION

In developed countries, ADR is the fourth leading cause of the death due to poor reporting of side effect after consuming drugs. The physicians sometimes can unknowingly prescribe excess dose and unwarranted combination of drug. In a country like India, ADR is scarcely reported to the ADR monitoring Centre named as Pharmacovigilance Programme of India (PvPI). Reporting of ADR to PvPI is slow because half the population of patients depend on the drug stores and self-medication. In addition, the physician who hold alternative medicine qualifications sometime prescribe the allopathic drugs also.

The pharmacovigilance educate the public about the benefits and risks of medicines. Patient should be made aware about the clinical trials and post-marketing surveillance through the statistical report of adverse drug

reaction available in websites such as FDA, CDC and PvPI. The PvPI aims for the best use of drug to cure the disease with no adverse effects. The PvPI works in the process of pre-marketing animal studies and post-marketing surveillance to reduce the adverse effects of new drugs on humans. Initially, the drugs are tested on animal but there is limitation of testing of drugs in animal which may not predict human toxicity. Post-marketing surveillance done after the drug is released in the market monitors the drug through data collected from reporting database, record linkage, electronic health records, prescription, event monitoring and patient registries. These data are processed using data mining to alter or confirm the drugs for further usage.

The data collection for ADR need validation due to the demography. For example, the data collection vary with the literacy rate, consumer behavior knowledge, and attitude. Pharmacovigilance relies on well acknowledged and obvious quality management systems that will benefit not only those involved in data collection, management and output production, but the patients who are the pharmacovigilance end users.

In this paper, the Natural Language Processing – Artificial Neural Network (NLP-ANN) technique is proposed for analyzing the created data set in a particular demography. The created data set is the collection of the reviews obtained from the online website and offline data through the questionnaires from patient, nurse, and physicians. Furthermore, the NLP-ANN uses the sentiment analysis. This paper is organized as follows: section 2 describes the background work; section 3 explains the proposed methodology; section 4 the experimental results and section 5 the conclusions.

II. RELATED WORK

Adverse drug reactions (ADRs) deals with genuine health issue everywhere throughout the world, and how to identify ADRs in an early stage has drawn many scientists' attention. There is a sudden surge of studies concentrated in this area and numerous strategies have been proposed to recognize ADRs as human life is at stake[1]. In any case, these information sources are restricted by high cost and protection issues. More than 2 million genuine symptoms, including 100,000 events, happen because of adverse drug reactions (ADR) consistently in the US. In spite of the fact that different NGOs screen ADRs through self-reporting systems, prior identification can be accomplished utilizing patient Electronic Health Record (EHR) data accessible at numerous restorative offices. [1] The pharmacovigilance action based EMR framework will be savvy technique to recognize ADR signals

Paper[2] analysed the adverse reactions due to orlistat and the results are analyzed. In[3] impact of adverse effects due to Amiodarone and Dronedarone was analyzed A review was conducted in a private healing facility of São Paulo State, Brazil. All patients were matured ≥ 60 years, present in the general practice ward in May 2006, discussed about the drugs utilized and the manifestations/grievances that brought about hospitalization. Alarming more than a half (54.5 %) of elderly hospitalizations were connected with ADR[4]. The ADR were spread in circulatory (28.5 %), stomach related (20.0 %) and respiratory (6%) tracts. Polypharmacy was another ADR hazard [5] where the fault mostly lies with patients.

Premarketing investigations of drugs identify normal adverse events but cannot dependably recognize an increasing rate of uncommon adverse occasions[6]. Frameworks for evaluation of post marketing adverse events incorporate unconstrained reports, mechanized cases or therapeutic record databases and other post marketing blunders[7]. Post marketing surveillance is fundamental for building up a full understanding of the adverse events and its impacts

Underreporting is another genuine problem encountered in Pharmacovigilance[8]. Studies demonstrate that two essential reasons that cause the underreporting are: 1) healthcare experts are ignorant of experienced ADRs, particularly for those bizarre ADRs;

2) they are excessively occupied with, making it impossible to write full report. The authors in[9] addressed these two issues by building ADR reporting framework. The framework can 1) enable healthcare experts to recognize the causal connection between a drug and an ADR by breaking down patients' electronic records; 2) make the reporting significantly less demanding by naturally connecting the patients' electronic information with the reporting structure.

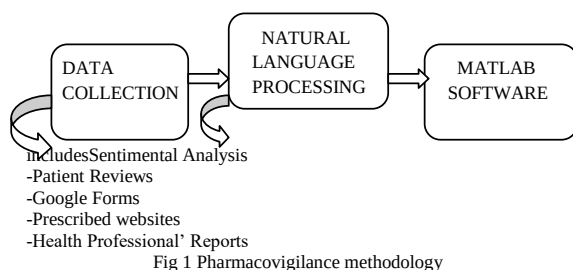
Randomized controlled trials (RCTs) are the fundamental premarketing techniques used to distinguish ADRs[10], for example, constrained review test size and restricted heterogeneity because of the rejection of the frailest patients. Post marketing drug safety checking through pharmacovigilance exercises, mining of details with more attributes gives scope for ADR identification and measurement thereby lessening healthcare costs [11]. The objective of the authors in [12] was to investigate the study of disease transmission and the costs of ADRs in routine care.

In [13] the authors proposed two novel calculations - a likelihood ratio and a Bayesian network for adverse drug effect discovery. In spite of the fact that the execution of these two calculations is practically identical, Bayesian certainty proliferation neural network, contain information on patients' visits, condition, treatment, and test outcomes. As adverse reactions to drugs are a main source of hospitalization and demise around the world, a novel Feature-based Similarity display (FS) to identify the potential ADRs for therapeutic cases utilizing the electronic patient dataset is also proposed. In [14], the authors have compared the adverse events of the drugs Acetaminophen and Ibuprofen.

Adverse Drug Reactions (ADRs) has turned into an overall issue that draws consideration of individuals from all racial and ethnic gatherings. Research discoveries demonstrate that generally ADRs can be effectively reduced to some degree by utilizing computer aided information advances[15]. In spite of the fact that numerous spontaneous reporting systems (SRSs) have worked to upgrade the pharmacovigilance, the ADRs data is still extremely inadequate to distinguish a causal connection between an adverse occasion and drug.

III. PROPOSED METHODOLOGY

The methodology of the pharmacovigilance is structured in an orderly manner as shown in Figure 1.



The data collection is done through the online and offline modes. In the online mode, the collections are done through the google forms, prescribed websites, and health professional's reports and in the offline mode the questionnaire forms are collected from the patients who used the drugs from the drug store and in hospitals. The google form consists of the simple online form, with the question on surveys which can be shared as a link, or embedded in a mail and website. All the response to the questions are neatly arranged in spread sheet. These forms are stored in google drive and can be accessed anywhere and from any device. Figure 2 shows the google form used for data collection.

DRUG REVIEW

Drug safety (Pharmacovigilance) is widely used for monitoring the adverse effects of the licensed drug. The review is based on good medication status by patient for analysis of pros and cons of drugs.

NAME: _____

AGE:

☐ 15-20 ☐ 21-40 ☐ 41-60

GENDER:

☐ Male ☐ Female

Figure 2. Google Drug Review form

The prescribed websites like Drugs.com and WebMD provide the information on drugs for consumers and healthcare professionals. Furthermore, the Drugs.com website also provides the consumer reviews after the intake of pills. WebMD site provides the health care tips; symptom checklist and pharmacy inform health professionals who reports action in the blog of physicians.

The reports submitted by consumers, health professional, manufacturers and distributors is considered as health professional reports. As per FDA, manufacturers and distributors should submit the report mandatorily. These data provide the conclusion on the safety of the health products. The reviews in the forms are analyzed using natural language processing.

Natural language processing will take string of words as the input and produce the structured representation as the output by capturing the meaning of those strings. NLP consists of number of stages as shown below in the Figure 3.

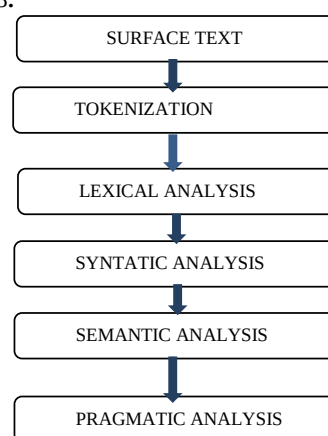


Figure 3 Block Diagram of NLP process

IV. EXPERIMENTAL RESULTS

The Figure 4 shows the negative analysis of sentiment analysis. The text mining determines the high-quality information from unstructured text collection from online posts such as twitter posts, scientific papers, or reviews, which depend on the organization to conduct the analysis. From the text analysis the answers obtained depend on what the customers discuss, talking and feel. The customer discussion obtain from comprehensive text analysis depend on contextual idea from the word to show the prediction of future behavior.

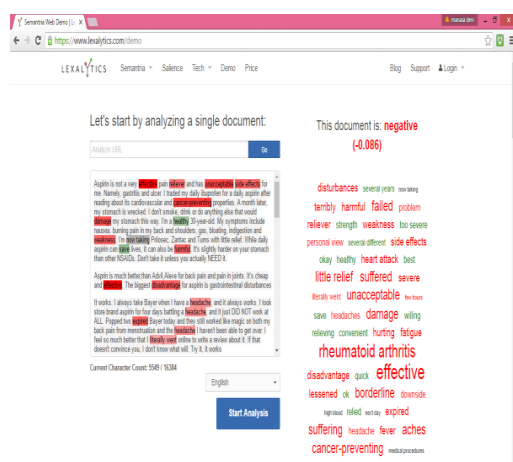


Figure 4 Negative analysis through sentiment analysis

In the Figure 4 the negative analysis of drug review is posted for the analysis. In the above analysis the paracetamol tablets reviews are posted and after posting the reviews in one website, the sentiment analysis provides the result as shown in Figure 5.

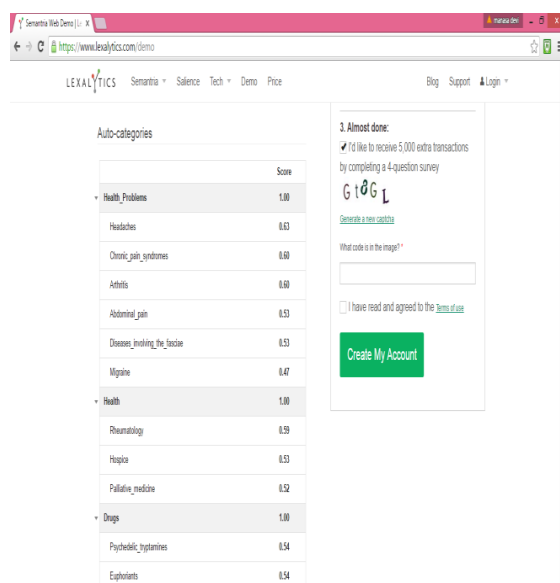


Figure 5 sentimental analysis of Paracetamol tablet.

In the sentiment analysis, the result shows exactly how the consumers feel about the drug. The sentiment analysis evaluates how strong the sentiment is, as it is a powerful tool, more accurate and reliable to provide whether the

text is positive, neutral, or negative. The rating of evaluation is assigned with score to conclude which sentiment is strongest. Once the score for the drug is obtained from different website, social media and direct collection of review from google form, it is analyzed with ANN using Matlab software. MATLAB is the high-level Matrix/Array language-programming tool, which performs simple and complex programming. The library can be written in C and FORTRAN and interfaced with MATLAB.

Figure 6 shows the Epochs of best validation. The Error Weights (EW) can have a size of 1, and is automatically extended to match the targets T. It allows for conveniently weighting the dimension and time, $TS=1$ and if all dimensions are 1, for instance if $EW = \{1\}$, then all target values are treated with the same importance.

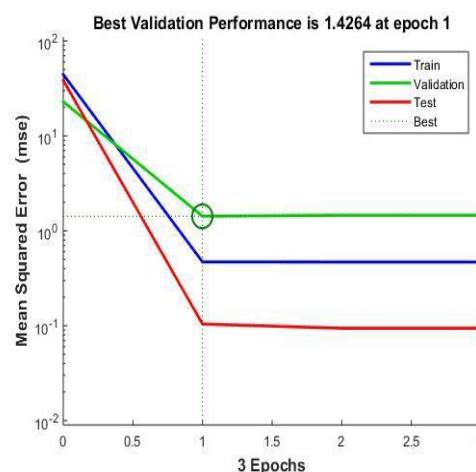


Figure 6 shows the Epochs for best validation

Regression analysis in MATLAB is a statistical process which provides the information on several variables to arrive at a relation between dependent variable and/or independent variable. MATLAB also uses the Feed Forward Neural Network to arrive at a statistical report. Figure 7 shows the regression analysis of website and social post, with an R value of 0.65 and proves that there is deviation in the post and reviews in the website.

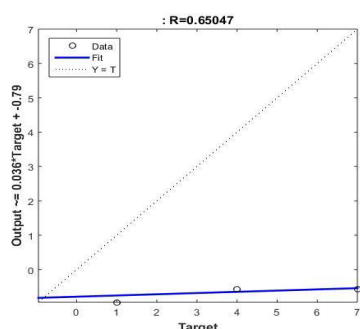


Fig.7 Regression analysis: social post/ drug website.

Figure 8 shows the regression analysis of the website and google form which shows the rating of the website to be more positive when compared to the social posts.

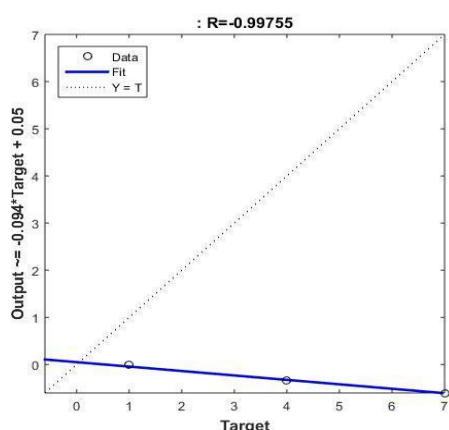


Figure 8 shows the regression analysis for the drug website and google form

Fig.8 shows that the regression analysis has negative R value which indicates that the comments posted and real-time data gathering varies in text mining.

V. CONCLUSION

The reviews collected from Googleform, prescribed websites, and health professional reports are processed for sentimental analysis. These processed outputs are analyzed in MATLAB and the regression value is calculated and represented in graph. This analysis provides the efficient way for rapid assessment of the drug toxicity in clinical data collection rather than social media

or website. Furthermore, the result shows that the data collected in the clinic proves to be more efficient for analysis of ADR.

REFERENCES:

- [1] Olsson S et al. An analysis of pharmacovigilance activities in 55 low and middle income countries. *Drug Safety*, 2010, 33(8):689–703
- [2] Poongodi, Latha Parthiban. Analysis of Adverse reaction associated with dietary drug Orlistat using data mining algorithms” in *International Journal of Pharma and Bio Sciences*, 2017 July; 8(3):(P)74-80, ISSN : 0975-6299.
- [3] G.Sriramakrishnan et al (2018) Identifying the Safety Profile of Amiodarone and Dronedaron using Openvigil, *International Journal of Pure and Applied Mathematics*, Volume 118 No. 18 2018, 2779-2784
- [4] Reyes-Morales , Mino-León D, Doubova SV, Flores-Hernández S. Postmarketing pharmacovigilance of adverse drug reactions: the case of rosiglitazone in Mexico. *Int J Clin Pharmacol Ther*. 2012, 50(1):1-9
- [5] Mara Lynn Wilson Prediabetes *Nursing Clinics of North America*, Volume 52, Issue 4, Pages 665-677, 2017
- [6] Levesque C. *Management of Diabetes in Children and Adolescents*. *Nurs Clin North Am*. 2017 Dec;52(4):589-603. doi: 10.1016/j.cnur.2017.07.008. Epub 2017
- [7] Pariente A, Didailler M, Avillach P, Miremont-Salame G, Fourier-Reglat A, Haramburu F, Moore N (2010) A potential competition bias in the detection of safety signals from spontaneous reporting databases. *Pharmacoepidemiol. Drug Saf* 19: 1166-1171.
- [8] Papay J, Yuen N, Powell G, Mockenhaupt M, Bogenrieder T (2012) Spontaneous adverse event reports of Stevens-Johnson syndrome/toxic epidermal necrolysis: detecting associations with medications. *Pharmacoepidemiol. Drug Saf* 21: 289-296.
- [9] Gregoire F, Pariente A, Fourier-Reglat A, Haramburu F, Begaud B, Moore N (2008) A signal of increased risk of hypoglycaemia with angiotensin receptor blockers caused by confounding. *Br. J Clin Pharmacol* 66: 142-145.
- [10] Wang HW, Hochberg AM, Pearson RK, Hauben M (2010) An experimental investigation of masking in the US FDA adverse event reporting system database. *Drug Saf* 33: 1117-1133.

- [11] Hauben M, Hochberg A (2008) The Importance of Reporting Negative Findings in Data Mining. The Example of Exenatide and Pancreatitis. *Pharm Med* 22: 215-219.
- [12] Jones SC, Sorbello A, Boucher RM (2011) Fluoroquinolone-associated myasthenia gravis exacerbation: evaluation of postmarketing reports from the US FDA adverse event reporting system and a literature review. *Drug Saf* 34: 839-847.
- [13] Piccinni C, Motola D, Marchesni G, oluzzi E (2011) Assessing the association of pioglitazone use and bladder cancer through drug adverse event reporting. *Diabetes Care* 34: 1369-1371.
- [14] Smita C Thomas et al (2018) Investigation of drug induced adverse effects and Disproportionality Analysis of Acetaminophen and Ibuprofen, *International Journal of Pure and Applied Mathematics*, Volume 118 No. 7 2018, 667-674
- [15] G. Sriramakrishnan et al (2016) Data mining techniques for finding serious Adverse Events *Journal of Chemical and Pharmaceutical Sciences* Vol 9, No 4, pp.1-5
- [16] R. Joseph Manoj, M.D. Anto Praveena, K. Vijayakumar, "An ACO-ANN based feature selection algorithm for big data", *Cluster Computing The Journal of Networks, Software Tools and Applications*, ISSN: 1386-7857 (Print), 1573-7543 (Online) DOI: 10.1007/s10586-018-2550-z, 2018.
- [17] K. Vijayakumar, C. Arun, Automated risk identification using NLP in cloud based development environments, *J Ambient Intell Human Computing*, DOI 10.1007/s12652-017-0503-7, Springer May 2017
- [18] K. Vijayakumar, C. Arun, Continuous security assessment of cloud based applications using distributed hashing algorithm in SDLC, *Cluster Computing* DOI 10.1007/s10586-017-1176-x, Sept 2017
- [19] K. Vijayakumar and C. Arun, "Continuous Security Assessment of Applications in Cloud Environment", *International Journal of Control Theory and Applications*, ISSN: 0974-5645 volume No. 9(36), Sep 2016, Page No. 533-541.

