Abstract: Due to growth of science and technology, the amount of unstructured health care data produced in social media in internet world is growing in rapid manner compared to structured health care data. This leads to quiet vast range of data availability in social media for analysing Adverse Drug Reactions of medicines, termed as Pharmacovigilance. The main objective of Pharmacovigilance is to detect, assess, and understand the adverse effect due to intake of medicines and prevention of these reactions. The dynamic growth in the unstructured social media data presents both challenges and opportunities in analysis which leads to requirement of novel solutions rather than using traditional analytic tools. Hence, Big data plays an important role in recent research works done on pharmacovigilance as it provides more efficiency in scalability. In this paper we did literature review of the research work done on Adverse Drug Reactions using social media data implemented in Big Data and Hadoop.

Keywords: Data mining, Pharmacovigilance, Big data, Hadoop, ADR, Social media.

I. INTRODUCTION

A. Pharmacovigilance

Drug Safety Surveillance is the pharmacologic science identifying with the accumulation, identification, evaluation, checking, and avoidance of unfavourable impacts with pharmaceutical products [1]. This is an imperative process that permits wellbeing administrative powers to keep on evaluating advantages and chances all through the life-cycle of a medication and possibly identify genuine unfriendly occasions and new medication wellbeing flags that were most certainly not recognized before promoting approval. The procedure by and large contains data got from patients, social insurance suppliers, medicinal writing, doctors, pharmaceutical organization's deals group or drug specialists. Data gathered from various aforementioned sources ought to be prepared in a characterized predictable manner for electronic accommodation to the administrative powers such as FDA (Food and Drug Authority), WHO (World Health Association), MHRA (Medicines and Health Regulatory Agency), EMA (European Medicines Agency) and other nearby powers [2]. Aside from administrative necessities, pharmaceutical organizations need to accomplish this to serve open wellbeing, and to encourage a feeling of trust in patients in the drugs they utilize, and in conclusion yet above all, to proactively screen drug impacts to keep item withdrawal from business sector because of safety issues.PV is required for deliberately distinguishing causal affiliations in the between medications and symptoms and taking restorative activities, both for new medications being propelled, and additionally for medications as of now being used. It depends on the gathering of suddenly reported Adverse Event (AE) reports. Voluntary health professionals and consumers initiate Report [1].

B. Big Data

Considering the volume, complexity and growth rate of social media data, Big data is the right choice as in implementation component as it has got capabilities to capture, process and manage these kind of data efficiently in faster manner compared to our conventional technologies and tools, that too within in the expected time frame for usage. Social network is growing in terms of users tremendously in this era. Huge numbers of people participates in social medical forums and add their experience, effects and comments against their medication and treatment. Unstructured data generated from these forums could be loaded and processed to get the Healthcare output using big data which would be used to monitor and manage ADR by pharmaceutical companies. The data set of size ranging from 30 to 50 terabytes to multiple petabytes is considered as ‘big data’. Figure 1.0 provides layered architecture of big data system which has got three layers, including Infrastructure, Computing and Application Layers [3]. Big data could be described in terms of ‘V’s: Volume, Velocity, Variety, Veracity and Value. Quantity of data is referred as Volume, rate at which the data is received is referred as ‘Velocity’, complexity of data is related to Variety and correctness, quality of data relates to Value [14].

Figure 1.0 – Layered Big data architecture
C. Hadoop

Hadoop is implementation framework for Bigdata which is developed in Java. This uses MapReduce programming technique in a distributed computing environment. Google developed hadoop with MapReduce concept. Hadoop is a platform that provides both distributed storage and computational capabilities. Apache hadoop system has got Hadoop kernel, MapReduce, HDFS and number of various components like Apache Hive, Base and Zookeeper. Hadoop provides the capacity of running business applications with high volume data without interruption by having more than one data nodes in transferring data in a faster manner. [3].

Hadoop is a distributed master–slave architecture that consists of the Hadoop Distributed File System (HDFS) for storage and the MapReduce programming framework for computational capabilities. The HDFS stores data on the computing nodes providing a very high aggregate bandwidth across the cluster. Traits inherent to hadoop are data partitioning and parallel computation of large datasets. The scalability is high in terms of storage and data processing with ability of adding additional resources like data node. [3].

Hadoop has got fault tolerance storage system named Hadoop Distributed File System (HDFS) which could store huge amount of data using multiple data nodes and survive data failure in data storage by storing data as multiple b locks in multiple nodes. Below Figure describes the Hadoop architecture. Hadoop will continue to work with alternate node available if failure happens in one node so that user will not get interrupted. Data would be replicated on each data node to maintain the data availability in case of failure. [3]

D. Social media

Social networks concentrating on health related topics have seen rapid development in recent years. Users in an online community often share a wide range of their medical experiences. These social networks connecting the people worldwide and allowing them share similar conditions—something that might not be possible in the real world. [7]. Social media plays vital role in providing a platform for patients to share their drug usage of experience. The on-line discussion forums, social networking sites, twitter feeds, blogs and search query results are main social media sources for ADR analysis. Monitoring the online resources like these would benefit stake holders of health care industries to have an evidence of ADRs. [8] Users discuss their health issues in the forum like Dailstrength[5], MedHelp[6] which contains prescription of drugs, side effects and treatment. ADR monitoring mainly focused on using data from these online forums. The Pew Research Center’s overview, The Social Life of Health Information [15], found that 2% of patients and 6% of caregivers share their experiences on the web, and that 18% of all internet group, 31% of all patients with ever ending conditions, and 38% of caregivers look at online drug surveys. This rapid growth of social media is giving new opportunities for public health surveillance that are internet-based, patient-generated, unsolicited, and up-to-date. [9]. It has become a challenge for research community in getting the relationship among drugs and their impacts from data provided by internet community in social web forums. [8]. Asscertained the significance of processing huge volume of medical data with Hadoop and MapReduce. They concluded that Hadoop is best choice to handle clinical data at lower cost in best performance. [4]. Public database, Biometrics, Bio medical images and Bio medical signal and the challenges in handling them.

Dr.Smitha Rao et al illustrated necessity of secured big data analytics for Healthcare industry. They mainly focused the
research on security issues, solutions when big data technology is used on health care data. The paper provides an insight to regulatory and technical challenges in handling health care data. It discusses the reasons for the security breach compared to traditional technologies. They proposed four security models named data de-identification, data centric approach to security, walled garden model and jujutsu security [10].

Emad A Mohammed Et Al enhanced us in Big data implementation using MapReduce programming for Clinical data. Different type of analytics has been discussed and MapReduce technique is proposed with Hadoop platform out of all these kinds. They discussed all kinds of medical data like Ming Yang et al proposed an automated solution to filter the consumer ADR messages from Example user posts in Social Media

![Example Posts](image)

II. LITERATURE REVIEW

F. Martin et al explored and analysed the significant challenges of data management in bio Big data in medicine. Authors described structured and unstructured big data in terms of Health and Bio medicine. This paper highlighted the importance of big data by notable changes that are happening in clinical and biomedical research. Also, they discussed the social media data usage in doing Adverse Drug Reactions from internet resources like Google search log, on-line news sources and Twitter[8].

Social media and they filled technical challenges in current ADR detection. They used classification method to filter social media especially for selected drugs. They analysed the background of using text mining on social media data and classification for labelling the training data. Four methods named EAT, PNLH, ACTC and Laplacian SVM are compared with proposed approach and comparison results are tabulated for showing the performance of the proposed approach [11].

Harshawardhan S. Bhosale et.al did review on recent research work done on Big data and hadoop. They discussed the concept of Big Data with its features, Hadoop framework and MapReduce architecture. They did extended comparison of various hadoop components and tabulated the same [3].

Aheed Sarker et al in their research, they did extensive survey on the analysis of Adverse Drug reaction from the social media data. They performed systematic approach for classifying the different methods to get the ADR detection done from social media. Also they studied on applications of this to pharmacovigilance. They analysed in the aspect of advantages and disadvantages of monitoring social media for ADR. This paper summarises the comparison and review of the research work done using social made for about past 4 years and tabulated with relevant details. Both quantitative and qualitative evaluation methods are discussed. A framework for ADR detection and extraction from Social media is proposed. This indicates the social media data become one of the critical sources for doing ADR analysis [12].

M.K.Ross et al have done a survey on “big data” at the sense of Electronic Health Records.

They focused on the research work on EHR from PubMed for the period of two years from 2011 to 2013. Three different query terms for the categories named EHR,Data Analysis and Health information technology. This paper also discussed the articles in three angles, Data mining, Pharmacovigilance, NLP and Phenotyping. Social media data utilization for ADR also been discussed along with personal monitoring [2].

Tao Huang et al in their research paper, discussed and summarised the applications of Big data to social medical data and recommendation systems in health care. Also the techniques of applying Big data to epidemic surveillance and eQTL are discussed. Over view of Hadoop components like HDFS and MapReduce are discussed [13].

III. CONCLUSION

Our survey covers research efforts for ADR analysis using Social media data limited to the papers those have implementation with Big data and Hadoop. We could find very few contributions relevant to this area, particularly implementation using Big data and Hadoop. Our review results shows that there is a wide opportunity in doing the research on pharmacovigilance on social media data implemented in Hadoop stack as it has got high performance capability in storing, retrieving and managing. We could focus on the social media data from mobile devices from applications like whats app, hike and vibher and sms for analysing ADR which is not explored and used for research as of now for ADR.

References


