Detecting Adverse Drug Reaction from Social Media Data using Deep Learning Approach

By

Sanam Aamir

(Registration No.: 0000321010)

Thesis Supervisor: Dr. Muhammad Usman Akram

DEPARTMENT OF COMPUTER & SOFTWARE ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

JULY 2023

Detecting Adverse Drug Reaction from Social Media Data using Deep Learning Approach

Author

Sanam Aamir

MS19CSE(SE)00000321010

A thesis submitted in partial fulfillment of the requirements for the degree of

MS Software Engineering

Thesis Supervisor

Dr. Muhammad Usman Akram

Thesis Supervisor's Signature: Usman

DEPARTMENT OF COMPUTER & SOFTWARE ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY, ISLAMABAD

JULY, 2023

Annex A

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by **NS Sanam Aamir** Registration No. 00000321010, of College of E&ME has been vetted by undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the thesis.

Signature : Usman

Name of Supervisor: **Dr Muhammad Usman Akram**

Date: 10-07-2023

Signature of HOD:
(Dr Usman Qamar)
Date: 10-07-2023

Signature of Dean: Rashid
(Brig Dr Nasir Rashid)
Date: 10 JUL 2023

Declaration

I certify that this research work titled "*Detecting Adverse Drug Reaction from Social Media Data using Deep Learning Approach"* is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources has been properly acknowledged/ referred.

Signature of Student

Sanam Aamir

MS19CSE(SE)00000321010

Language Correctness Certificate

This thesis has been read by an English expert and is free of typing, syntax, semantic, grammatical and spelling mistakes. Thesis is also according to the format given by the university.

Signature of Student

Sanam Aamir

MS19CSE(SE)00000321010

Signature of Supervisor

Dr. Muhammad Usman Akram

Copyright Statement

Acknowledgments

All praise and glory to Almighty Allah (the most merciful, the most beneficent) who gave me the courage, patience, knowledge and ability to carry out this work and to persevere and complete it satisfactorily. Undoubtedly, HE eased my way and without HIS blessings I can achieve nothing.

I would like to express my earnest appreciation to my advisor Dr. Muhammad Usman Akram for boosting my morale and for his continual support, motivation, dedication and invaluable guidance in my quest for knowledge. I am blessed to have such a co-operative and kind mentor for my research.

Along with my advisor, I would like to acknowledge my entire thesis committee for their cooperation and prudent suggestions.

My acknowledgment would be incomplete without thanking the greatest source of my strength, my family. I am abundantly thankful to my beloved parents who raised me when I was not able of walking and proceeded to support me throughout in every department of my life.

Finally, I would like to express my gratitude to all my friends and the individuals who have encouraged and supported me through this entire period.

Abstract

Adverse Drug Reactions (ADRs) are very common and cause serious consequences to patients. Detecting them can be a very difficult task. With the increasing popularity of social media platforms, they have become a hub of data. A lot of data related to identifying potential ADRs can be found on social media. But extracting useful information from it can be a challenging task as the data is in unstructured form and has a sheer volume. This study proposes an approach to detect and list unknown ADRs from social media data using machine learning and NLP based techniques. The framework utilizes Natural Language Processing (NLP) to automate the discovery of ADRs mentioned in social media posts. They are then compared to a list of known ADRs to identify unknown ADRs. The dataset for this study has been self-collected and contains tweets related to ADRs. Three drugs were shortlisted for this study; Adderall, Xanax, and Prozac. For Adderall and Xanax, one unknown ADR each was found, whereas, for Prozac, three unknown ADRs were found. The proposed approach can be used to cater to different problems in addition to identifying unknown ADRs in the future. This study improves patient safety by providing a new approach to detect unknown ADRs from tweets, contributing to the field of pharmacovigilance.

Keywords - *Adverse Drug Reactions (ADRs), Social Media, Natural Language Processing (NLP), Word Embeddings, Word2Vec Model, Cosine Similarity*

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1. Adverse Drug Reactions

Sometimes, when a patient is given medication, it can cause some harmful and undesirable side effects. These are called ADRs. They can be some mild side effects like nausea or dizziness, or some serious side effects like disabilities, organ failure, or even death. ADRs should not be ignored as they can affect the patient's treatment outcomes, health, or quality of life.

If we look at the statistics, it can be seen that ADRs are a significant public health concern around the world. Every year, millions of people are hospitalized or even die as a result of ADRs. According to World Health Organization (WHO), ADRs are the fourth most common cause of death in the USA. Every year, it affects around 1.5 million people in the USA alone. In Europe, around 5% of the admissions in hospitals are because of ADRs, and in the United Kingdom, this percentage is 6.5%. Moreover, in the past two years, the rate of death because of ADRs has increased by 7.5%, whereas the rate of severe reactions has increased by 18% [1]. During a recent study by Al-Shareef *et al.* [2], it was seen that 56% to 78% of COVID-19 patients had a prevalence of ADRs. Figure 1.1 shows the type and occurrence of ADRs in the patients of that study.

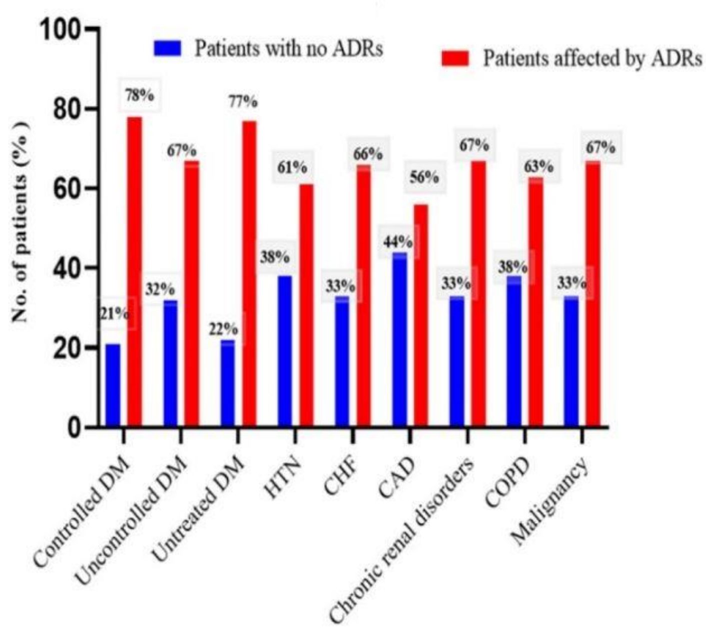

Figure 1.1 Type and Occurrence of ADRs in COVID-19 Patients [2]

Disease severity, medication type, and patient demographics also have an impact on the occurrence of ADRs. A patient using medications including chemotherapy drugs, anticoagulants, and antibiotics is at a higher risk of developing ADRs. The severity of ADRs is also dependent on different factors like the duration of the treatment, the dosage of the medicine, the presence of other drugs, and patient's health status and age.

In order to reduce or alleviate the risk of developing ADRs, healthcare professionals take necessary steps to identify, manage, and prevent them. These steps may include educating patients about potential side effects and medication adherence, adjusting medication doses, and monitoring patients closely for signs of ADRs. Patients should also play their part in controlling ADRs by reporting any such incidence to the healthcare professional immediately.

1.2. Pharmacovigilance Systems

Pharmacovigilance systems are activities and processes designed for reporting, evaluating, and monitoring different drug-related problems including ADRs. Despite the numerous lifesaving benefits of medicines, the use of medicines has some risk associated with it. These risks are identified and minimized by using pharmacovigilance systems.

Various different components make up the pharmacovigilance system. These components are used to perform different task like to access the causality of ADRs, to detect potential safety issues, to analyze and collect information related to ADRs, to implement risk management strategies for minimizing or preventing harm to patients.

Pharmacovigilance systems get their data from different sources like electronic health records, observational studies, clinical trials, and spontaneous reports from patients and doctors. Potential health and safety concerns are identified from these sources. They also serve as an evidence in favor of or against a drug. But there are some challenges being faced by pharmacovigilance systems. ADRs are underreported through conventional reporting methods. The reasons behind it are the complex reporting process, misconception that ADRs are not serious, and lack of awareness. There is a need for identifying a source that reports ADRs spontaneously.

1.3. ADRs and Social Media

With the increasing popularity of social media platforms, it has become a hub of information. It is observed that ADRs are also being discussed there. Many experiences of patients and doctors with medicines can be found on social media posts, including Facebook posts, Tweets, Reddit posts, etc. these posts can be used as source to extract frequency, severity, and symptoms of ADRs.

Social media can serve as a real-time source of information for pharmacovigilance systems providing it with data reported by patients about ADRs. But the data on social media is in unstructured form and has a sheer volume. In order to extract useful information from it, NLP techniques and social media listening tools can be used.

1.4. Natural Language Processing

NLP is a field of study that falls under artificial intelligence and computer science. This field is related to the communication between computers and humans using natural languages. Under this domain, different algorithms are developed that interpret, generate, and understand human languages. The NLP pipeline consists of seven steps including sentence segmentation, word tokenization, stemming, lemmatization, stop word analysis, dependency parsing, and part-of-speech tagging. Figure 1.2 shows this pipeline.

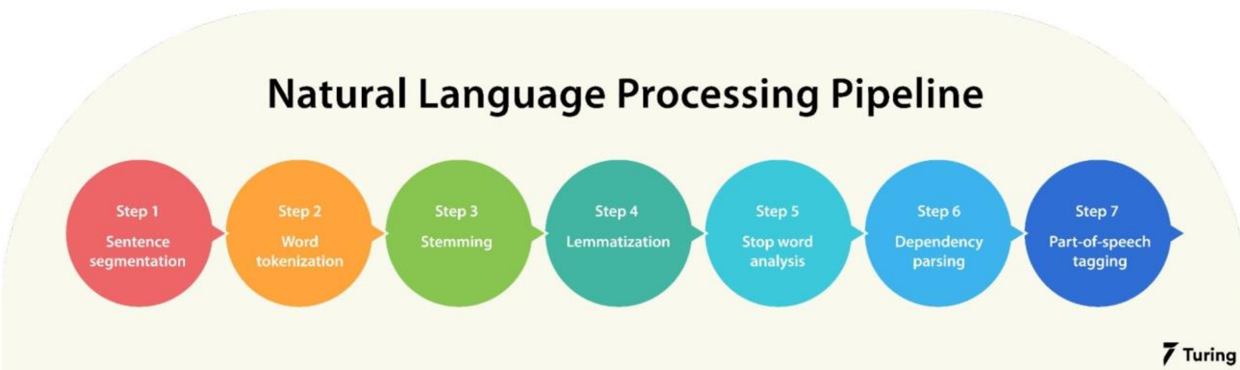

Figure 1.2 Natural Language Processing Pipeline [3]

In NLP, the data related to language is analyzed and processed using linguistic rules, statistical models, and machine learning techniques. Different insights related to the data are generated, questions are answered, and other tasks are performed using relationships and patterns in text. These relationships and patterns are recognized using linguistic rules, statistical models, and machine learning techniques.

The unstructured and sheer volume of data on social media platforms like Facebook, Twitter, and Instagram can be analyzed and interpreted using NLP techniques. These techniques have proven to be powerful tools for extraction of information.

There are numerous applications of NLP for social media analysis. One of them is sentiment analysis. In this type of analysis, different algorithms are used for the identification and classification of the emotional tones used in social media posts. This means that's the posts are analyzed whether they are in a negative, positive, or neutral tone. This analysis can help businesses in monitoring social media, managing crisis, and understanding the sentiments of the customers about their products or brand.

Another application of NLP for social media analysis is entity recognition. In this type of analysis, different algorithms are used for the identification and extraction of specific pieces of information like organizations, locations, names, etc. from unstructured data. This analysis can help businesses in tracking down how many times or where the brand or company has been mentioned. It can also help in monitoring specific keywords or topics of interest.

Other applications of NLP for social media analysis include topic modeling, which is a technique that identifies themes and patterns in large textual data, and language translation, which is a technique for analyzing text in different languages.

However, there are many challenges while using social media data because it is unstructured in nature and there is a frequent use of emojis, slang, and other non-standard language. To overcome this, extensive data cleaning and preparation is required.

1.4.1. Word Embeddings

Many applications of NLP like question-answering systems, sentiment analysis, text classification, and machine translation use word embedding technique for their implementation. Word embeddings convert textual data into numerical vectors in a high-dimensional space. These numerical vectors capture the semantic and syntactic meanings of the words in a specific language corpus. In this way, the words that have similar usage patterns and meanings are closer to each other as compared to words having different usage patterns and meanings.

A well trained machine learning model is used to generate these word embeddings. For training purpose, a large corpus of textual data is required. The probability of a word appearing close to another specific word is learned and predicted by this model. In this way, the learned associations between words are stored in the numerical vectors. The co-occurrence and patterns in data are used for this purpose.

1.5. Motivation

This study seeks to improve knowledge and awareness of the medications and their side effects among doctors and the general public, and help doctors in prescribing safer drug combinations using Machine Learning and NLP techniques. It aims at providing a framework that understands the side effects of approved drugs that may remain unreported. Incorporating these findings into health policies could increase efficacy and compliance. Undesired side effects can be reported to the healthcare provider or directly to FDA. Data from social media can be used to identify a drug's new side effects and this is a great step towards advancements in pharmacovigilance. Social media can reduce the cost of clinical trials and can aid to gather a drug's unknown side effects along with their known ones.

1.6. Problem Statement

Traditionally, doctors have recommended medications to their patients based on their common uses. For instance, Disprin is a drug typically used to treat headaches and flu, but after 35 years in clinical trials, it was shown that Disprin may also be used to treat hypertension patients' blood thinning needs. Similar to this, patients occasionally claim to have unique side effects that are uncommon and beyond the norm.

The noise caused by misspellings and slang in social media data is a significant concern. Additionally, medications are sold all over the world under several brand names. Sometime after taking a specific medicine, people will post on Facebook or tweet about their adverse effects or symptoms. The major goal was to obtain information from various social media sites on the negative effects of drugs. Some people experience unidentified side effects of the medicines and share them on social media. As more individuals are using social media to share their drug-related experiences and side effects, a model was required to link general words, treatments, and symptoms with corresponding diseases, utilizing keyword filtering techniques.

The major goal is to gather information from various social media sites on the negative side effects of advertised drugs using data collected from Twitter using Deep Learning and NLP techniques. Data from social media can help in identifying new side effects of a drug, particularly those side effects that are unknown and this is a great step towards advancements in pharmacovigilance. The analysis will provide insights to real-world problems that a drug user may encounter. Incorporating these findings into health programs might increase efficacy and efficiency. Undesired side effects can be reported to the healthcare provider or directly to FDA.

1.7. Aims and Objective

Following are the aims and objectives of this study:

- To develop an approach to determine unreported side effects of drugs using machine learning and NLP techniques
- To collect and analyze information from social media sites and perform effective pre-processing to determine pharmacological side effects

1.8. Structure of Thesis

The report is structured as follows:

Chapter 1 gives an introduction about the proposed topic, aims, objectives and motivation.

Chapter 2 covers the detailed literature review of existing methods and models adopted to perform data extraction using NLP techniques.

Chapter 3 gives the materials, including dataset and tools used for implementation.

Chapter 4 discusses the proposed preprocessing steps and architectural details of the proposed model.

Chapter 5 discusses the experimentation including the setup used for implementation, quantitative results obtained, their discussion and benefits in comparison to other approaches.

Chapter 6 concludes the topic by suggesting some future work that is not under the scope of this research but can be implemented in future.

Chapter 2: Literature Review

This chapter reviews the current state-of-the-art models and approaches in a comprehensive but critical manner to find out their shortcomings. The literature reviewed below focuses on different deep learning and NLP methods and the benefits and downsides of using social media data to identify unknown adverse drug responses.

2.1. Deep Learning for Text Classification

Deep Learning is a subfield of Machine Learning that deals with artificial neural networks and their ability to learn from data. Text classification is one of the most common applications of Deep Learning. Text classification involves categorizing text into predefined categories, such as spam vs. non-spam emails or positive vs. negative reviews.

The Convolutional Neural Network (CNN) is a popular text categorization architecture. In CNNs, features are extracted from the text using a number of convolutional layers, and these features are then passed to a fully connected layer for classification.

The Recurrent Neural Network (RNN), which is made to manage sequential data like text, is another well-known model. RNNs can be taught to categorize the text based on the complete sequence and use a hidden state to retain context across the words in the sequence.

Models like the Transformer architecture have produced encouraging results in text classification tasks in recent years. The Transformer model has attained state-of-the-art success on many natural language processing tasks, including text categorization, by using a self-attention mechanism to understand the relationships between words in the text. The Transformer model's architecture is depicted in Figure 2.1.

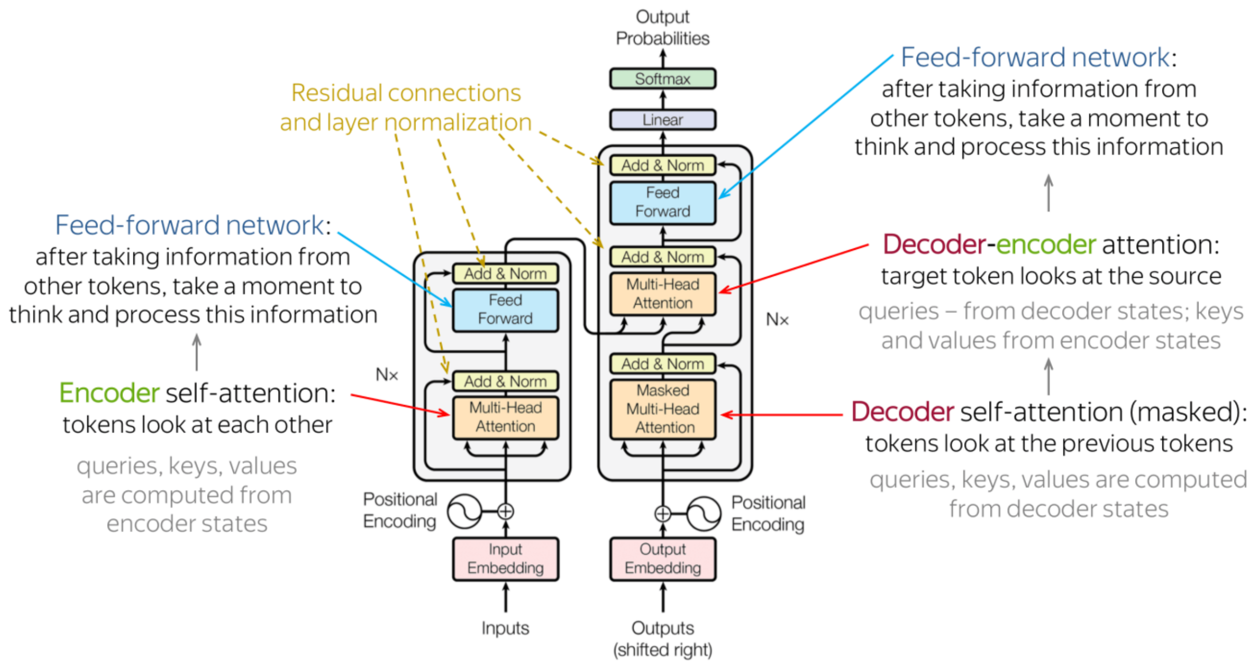

Figure 2.1 Architecture of the Transformer Model

Zhang and Wallace [4] gave a thorough study of CNNs for sentence classification, as well as sensitivity analysis and a practitioner's guidance. The authors assessed the efficacy of CNNs using four benchmark datasets. They discovered that CNNs regularly outperformed other models on all datasets, with three of them reaching state-of-the-art results. The sensitivity analysis showed that the number of filters and filter sizes have the greatest influence on CNN performance, while the activation function and dropout rate have a more modest impact. The writers also discovered that pre-training word embeddings on a big corpus can enhance CNN performance even further.

Shen *et al.* [5] proposed a CNN-based approach for learning semantic representations for web search, which can also be applied to text classification tasks. They represented documents and queries as fixed-length vectors. These vectors could be ranked and compared with each other easily. A pre-trained model was used to tokenize and encode documents and queries into word embeddings as the first step. For the next step, a multi-layered CNN was used for extracting features from these word embeddings. These features consisted of local and global information about the text. They were then used to generate fixed length vectors. Two publicly available large scale datasets were used for training and testing the approach. The results were compared with other state-of-the-art models as well and their approach outperformed.

Another study [6] used deep learning techniques to design an ensemble framework for text classification. The ensemble combined different models including Long Short-Term Memory (LSTM), Bidirectional Encoder Representations from Transformers (BERT), and CNNs. There were three modules in the proposed approach; the feature extraction module, the model training module, and the prediction module. In the first module, textual data was represented using pre-trained word embeddings. The second module trained the different models by iteratively setting their architectures and hyperparameters. The last module made the final predictions by combining outputs from different models to form an ensemble. The proposed approach outperformed other state-of-the-art models with an accuracy of 96.25%.

Liu, Li, and Hu [7] performed short text classification by designed their own approach. This approach also had three modules; the feature extraction module, the multi-stage attention network, and the classification module. Different experiments were conducted to test the efficiency of the modules. The results of these experiments showed that multi-stage attention network and context-relevant features improved the performance for short text classification.

2.1.1. NLP for Adverse Drug Reaction Detection

NLP techniques have become very common for ADR detection. Data is taken from pharmacovigilance databases, Electronic Health Records (EHRs), and social media. One of the reasons behind this is that NLP techniques can effectively and efficiently process a large volume of unstructured data. Apart from text, NLP techniques can also extract unstructured data from images and videos and convert them into structured form so that they can be used by machine learning models.

If we look at ADR detection, NLP techniques are being used to extract useful information from different data sources. This information may include patient characteristics, adverse events, drug names, etc. The techniques being used are topic modeling, sentiment analysis, relation extraction, and named entity recognition.

Many studies can be found in the literature that use NLP techniques for ADR detection. Grouin *et al.* [8] used clinical narratives in French language to detect and classify ADRs. A hybrid deep learning model consisting of RNNs and CNNs was designed for this purpose. This model captured and learned the textual characteristics of ADRs. Their model outperformed other conventional machine learning models in terms of accuracy and efficiency.

Zhang *et al.* [9] used EHRs as a data source for extracting information to identify and categorize ADRs. They combined CNNs with attention mechanism, developing a unique attention-based loss function. This loss function enhanced the model's interpretability. Their model proved to be a very efficient model for ADR detection and severity classification. Gaur *et al.* [10] proposed a new approach for detecting ADRs using RNNs and CNNs. Their source of data was social media. In order to improve the effectiveness of their approach, they designed a novel domain-specific word embedding model. Their model proved to be very efficient.

Tiftikci *et al.* [11] used rule based system and deep learning algorithm to identify ADRs in the text of drug labels. They used Medical Dictionary for Regulatory Activities (MedDRA) dictionary to normalize their data. They combined Conditional Random Fields (CRFs), CNN, and bi-directional Long Short-Term Memory (Bi-LSTM) for the identification of ADRs. Their in-house text-mining system, SciMiner, was extended using rule based system to normalize their data to MedDRA terms. Their approach showed promising results. Another approach was proposed by Zhang *et al.* [12]. They extracted text from EHRs and detected ADRs from them. The input text was automatically scanned and the most informative words and sentences were identified from it using hierarchical attention mechanism. They used a widely used benchmark dataset to test their approach. Their approach outperformed other state-of-the-art models. Li *et al.* [13] used an hybrid approach of NLP techniques to detect ADRs from clinical dataset. Their hybrid approach combined a machine learning technique and a rule based method. Their testing experimentation showed high recall and precision for their proposed approach on a large dataset.

Xu *et al.* [14] also proposed their novel approach to detect ADRs. They made the use of a graph neural network. This network modeled the relationships between diseases, symptoms, and drugs mentioned in the text. Their approach showed the best performance till then in biomedical literature. In another study by Hazarika *et al*. [15], social media data was used for detecting ADRs. The social media data was collected through Tweets from Twitter. They manually annotated the data. A combination of deep learning and rule based techniques was used to achieve this detection and showed high recall and precision in the testing phase. Liu *et al.* [16] proposed a framework that could simultaneously identify ADRs in EHRs, and categorize their severity level. Their framework showed good performance for multi-tasking.

The above literature showed that NLP techniques have proved to be very effective in detecting and classifying ADRs from unstructured text data sources. A great potential can be seen in them for improving patient safety through more accurate and faster detection of ADRs. They can identify potential ADRs that can’t be identified by traditional methods.

However, detecting and classifying ADRs from unstructured data using NLP techniques has some challenges as well. These challenges include potential bias in the data, a need for comprehensive and accurate ontologies, and the variation of completeness and quality of data across various data sources that can affect the overall performance of the models. Moreover, there is much room for improvement in terms of accuracy.

2.1.1.1. Word Embeddings for Adverse Drug Reaction Detection

Word embeddings are commonly used in text classification that is being done through deep learning models. Word embeddings is a technique that is used for mapping textual data to high-dimensional vectors. The semantic meanings of words, as well as the relationships between them are captured through word embeddings. Generally, a neural network is trained on a large dataset to create word embeddings. This neural network is trained to map textual data to high-dimensional vectors where the distance between vectors is decided on the basis of the semantic similarity between words. The lesser the distance, the more similar they are.

Word embeddings are very popular in different applications of NLP including text classification, language translation, and sentiment analysis. The reason behind this is that word embeddings improve the model accuracy and enhance the understanding of the meaning of text. Some commonly used word embedding techniques include Word2Vec, FastText, and GloVe models. The architecture of a CBOW and skip-gram based Word2Vec model is illustrated in Figure 2.2.

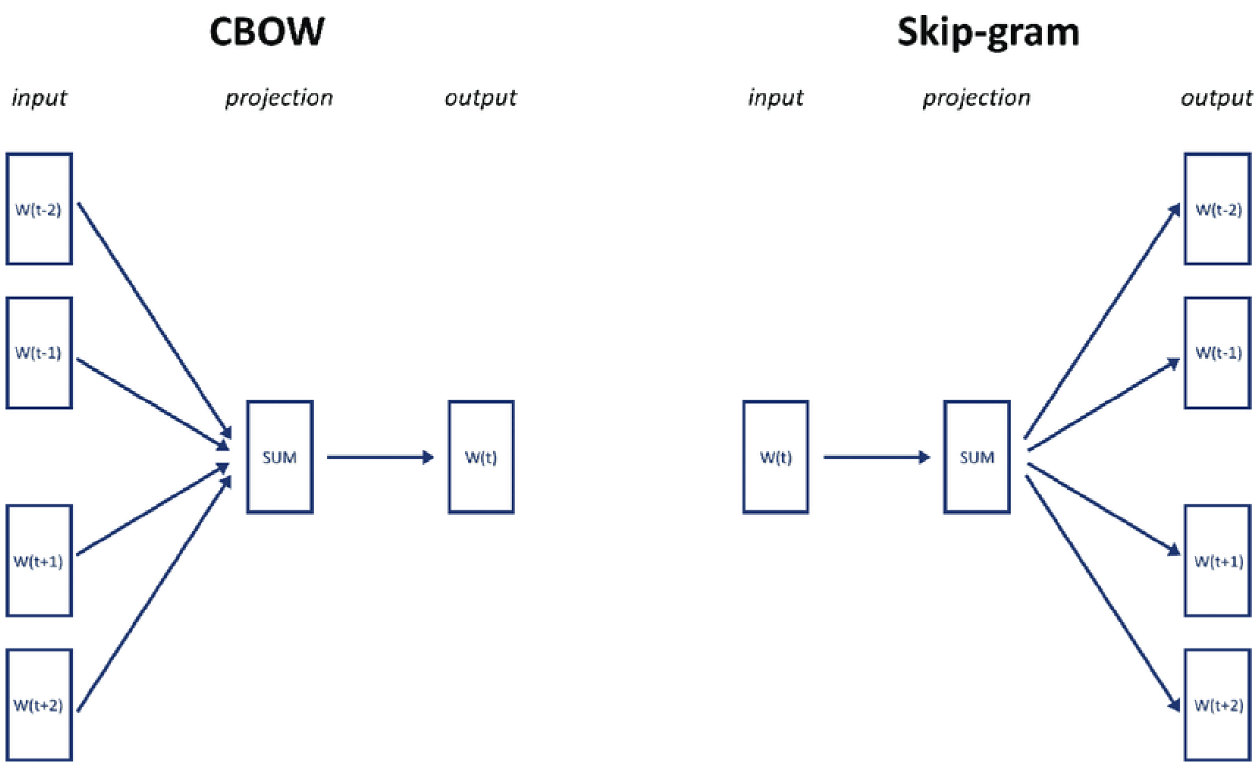

Figure 2.2 Architecture of a CBOW and Skip-gram based Word2Vec Model

Recently numerous studies have been conducted for investigating the use of word embeddings for the detection of ADRs from textual data. Below is a literature review of some of the studies focusing on this.

Miranda [17] proposed a new approach for the detection of ADRs using general purpose word embeddings and token-level convolutions. Using this approach, they found out that by removing redundant ADR relevant sentences, the overoptimism was greatly reduced from the classification results. Wu *et al.* [18] also successfully demonstrated that word embeddings used with CNN can produce a high accuracy for ADR detection. They designed their own model by combining the two techniques and used EHRs as a source of input data. Zhang *et al.* [19] designed their own variation of ADR detection framework from EHRs using hierarchical attention mechanism and word embeddings. Their framework outperformed many state-of-the-art models.

Tan *et al.* [20] proposed an innovative hybrid approach that combined machine learning and rule-based techniques to identify ADRs associated with different drugs. Rule-based technique was first used to identify the mention of an ADR within a distance of 100 characters from the mention of the drug. Next, machine learning technique was used to find the relatedness of that ADR with the drug. Another model was proposed by Li *et al.* [21] for detecting ADRs using a pre-trained language model called BERT and word embeddings technique. They got the data to train and test their model from social media platforms. The word embedding techniques that they used were FastText and GloVe models.

Masino *et al.* [22] developed a model for ADR detection. They performed this as a binary classification task where the tweets that they gave as input were classified as the ones containing mentions of ADRs, or the ones not containing mentions of ADRs. They used word embeddings to convert the textual data to vectors using unsupervised learning. The neural network they used was Convolutional Neural Network model (ConvNet).

Gao *et al.* [23] developed an NLP-based approach for ADR detection that uses a combination of word embeddings and attention mechanisms. The model achieved high precision and recall on a large-scale dataset of clinical notes.

Overall, these studies demonstrate the effectiveness of word embeddings for ADR detection in various types of text data, including clinical notes and electronic health records. These approaches can significantly improve the efficiency and accuracy of ADR detection, thereby contributing to drug safety and public health.

2.2. Adverse Drug Reaction Detection from Social Media Data

Social media has become a hub of various kinds of data. Data related to ADRs can also be found on social media in large quantities. Many studies can be found in literature that have used social media data to identify and detect ADRs. These studies have used different techniques including deep learning, machine learning, and rule based methods. Along with various pros, detecting ADRs from social media data has its cons as well. It is a challenging task. As the user-generated content is increasing day-by-day on social media, a growing interest can be seen in developing methods to detect ADRs from it. Following are some prominent studies that have used social media data to detect ADRs.

Zhang *et al.* [24] proposed a technique to identify ADRs using data from Twitter and DailyStrength. They used ADR lexicon and extended syntactic dependencies for extracting predicate ADR pairs. From each of these pairs, Part-Of-Speech (POS) and semantic features were extracted and a complete representation of deep linguistic features was generated by pooling all the features. Then they extracted several shallow features. Both these features were combined and given to the predictive model to train. Another study by Nikfarjam *et al.* [25] introduced a new machine learning based approach called ADRMine for detecting ADRs from user posts on social media. They used CRFs in their model. Multiple features were extracted and fed to the model including features to model similarities in words' semantics. The similarities were modelled using word embeddings.

Teklemariam *et al.* [26] developed an RNN based model for the labeling of ADRs in Twitter posts. The words from the Twitter posts were given as input to the RNN and it labeled them with ADR membership tags. The words were converted to numerical vectors through word embeddings before they were fed to the RNN. Bian *et al.* [27] designed an approach to identify potential ADRs related to some drugs by analyzing Twitter posts. They used a Support Vector Machine (SVM) classifier and NLP techniques to achieve this. The dataset for this study was very large scale. To overcome the challenges associated with the size of the dataset, MapReduce was used on a High Performance Computing (HPC) platform. Their approach showed promising results and suggested that social networking data could be used to identify potential ADRs that are an important patient safety issue. The authors of a recent study [28] presented a large corpus of annotated Tweets related to ADRs for the public to access freely. The tweets were collected using correct and misspelled versions of the names of drugs as keywords. They were then manually annotated to heck that whether an ADR was mentioned in them or not. They then trained SVM classifier and Naïve Bayes to automatically detect ADRs in those Tweets.

Ahmed *et al.* [29] introduced a novel approach for detecting ADR from social media data using a combination of convolutional and recurrent neural networks based on deep learning. Their method achieved the highest performance on a benchmark dataset of Twitter data, surpassing existing state-of-the-art approaches. Chen *et al.* [30] did a study on ADR detection from social media data and developed a hybrid approach that utilized domain-specific features, word embeddings, and a machine learning-based model. The study demonstrated high precision and recall in detecting ADRs from Twitter data. Table 2.1 shows a literature review of some more notable studies that have explored this topic.

Table 2.1 Literature Review of Some More Notable Studies that have Explored this Topic

Study	Methodology	Main Findings
[31]	Conversational agents and social media mining	Social media data can be effectively used to identify ADRs and provide timely and personalized advice to patients.
[32]	Multimodal deep learning	Multimodal deep learning can effectively detect adverse drug reactions from social media data by integrating text, image, and metadata information.
[33]	Deep learning model	Deep learning models can effectively detect adverse drug reactions from social media data by extracting features and patterns from textual data.
[34]	Deep learning-based natural language processing	Deep learning-based NLP can effectively detect adverse drug events from social media data by identifying relevant keywords and patterns.
[35]	Ensemble deep learning framework	An ensemble deep learning framework can improve the performance of adverse drug event detection from social media data by combining multiple deep learning models.
[36]	Deep learning approach	A deep learning approach can effectively detect adverse drug events from social

		media data by integrating multiple features and applying transfer learning.
[37]	Attention-based neural network	Attention-based neural networks can effectively detect ADRs from social media data by focusing on relevant keywords and phrases.
[38]	Graph convolutional neural network	Graph CNN can effectively detect ADRs from social media data by incorporating network structure and context information.
[39]	Ensemble deep learning approach	An ensemble deep learning approach can improve the performance of adverse drug event detection from social media data by combining multiple models and features.
[40]	Deep learning and network analysis	Deep learning and network analysis can effectively detect ADRs from social media data by considering the correlation between drugs, adverse events, and social media users.

Overall, these studies demonstrate the potential of social media data as a valuable source for ADR detection. The approaches used in these studies leverage different machine learning techniques and feature representations to improve the accuracy of ADR detection. These methods can provide timely and valuable information on drug safety to healthcare professionals and the public.

2.3. Gaps in Literature

After conducting a systematic review of existing research, and identifying inconsistencies in previous studies, some gaps in literature were identified. They are discussed below.

- **Limited Access to Social Media Data:** One of the major drawbacks in studies related to ADR detection from social media data is the limited access to high-quality social media data. Most social media platforms restrict the access to their data, and the publicly available data may not be representative of the entire population. Moreover, social media data may contain noise, which can affect the performance of the ADR detection models.
- **Lack of Annotated Social Media Data:** Another challenge in ADR detection from social media data is the difficulty in annotating the data. Moreover, the annotation process may require expertise in pharmacology and medical domain knowledge, which may not be available to all researchers.
- **Data pre-processing and noise removal are less explored:** Most studies related to ADR detection from social media data are focused on specific drugs, moreover, noise removal in the data is not catered. Unlike clinical data, social media data may contain informal language, sarcasm, and cultural references, which can be difficult to interpret. Moreover, the social media data may be biased towards certain demographics or geographies, which can affect the generalizability of the results.
- **Need for Human Validation:** While deep learning models can process large amounts of data quickly and efficiently, they may not always provide accurate results, especially when dealing with noisy and unstructured data such as social media posts. Human validation is needed to ensure that the results produced by the deep learning models are accurate and

reliable. This validation process involves manually reviewing the predicted ADRs and checking them against existing medical knowledge and literature.

2.4. Contributions of the Current Study

Based on an extensive review of literature and to overcome the limitations and challenges in detecting unknown ADRs from social media data, this study aims to utilize the Word2Vec and Cosine Similarity techniques to capture high-level accuracy. The proposed framework leverages NLP to automate the discovery of ADRs mentioned in social media posts. These will then be used to identify unknown ADRs. To achieve this, the study uses a new self-collected dataset containing a substantial number of sample social media posts from Twitter related to ADRs.

Utilizing the full text of social media posts gathered from Twitter, the proposed methodology identifies unknown ADRs that are mentioned in the posts using the Word2Vec and Cosine Similarity techniques. The study recognizes that a larger model size does not always guarantee greater performance on a cross-domain benchmark challenge.

Chapter 3: Materials

This chapter presents the work done for the collection of the dataset containing drug name and their side effects. The data primarily was extracted from Twitter. The drug names were used as keywords to search for the tweets. Tweets having the drugs' name and its side effect at the same time were collected. The data set of three drugs containing 149,568 valid data tweets were collected.

3.1. Drug Names Selection

To initiate the search, a list of drugs was compiled to be used as search keywords. Three drugs were identified, considering their widespread use and the availability of data on Twitter. The aim was to obtain a significant amount of information for the analysis. The drugs that were shortlisted were:

- Adderall
- Xanax
- Prozac

3.1.1. Adderall

What it Treats: Attention Deficit Hyperactivity Disorder (ADHD)

Uses: This combined medication is indicated for the treatment of ADHD. Its mechanism of action involves altering the levels of specific natural substances in the brain. Dextroamphetamine is classified as a stimulant drug.

Precautions: Misuse or improper use of Adderall may result in severe and potentially lethal cardiovascular and blood pressure complications. Medications containing amphetamines have the potential to induce addiction.

3.1.2. Xanax

What it Treats: Anxiety and panic disorders

Uses: Xanax is prescribed for the treatment of anxiety and panic disorders. This medication falls under the category of Benzodiazepines, which work by affecting the brain and nerves in the central nervous system, resulting in a soothing effect.

Precautions: It is important to note that using Xanax in combination with opioid medications (such as Codeine or Hydrocodone) may significantly raise the risk of severe side effects, including fatality. Therefore, it is essential to consult with a healthcare professional before taking any such medication.

3.1.3. Prozac

What it Treats: Depression, panic attacks, Obsessive Compulsive Disorder (OCD)

Uses: Prozac is a medication that is prescribed to manage various conditions, including depression, panic attacks, OCD, Bulimia, and Premenstrual Dysphoric Disorder (PMDD).

Precautions: Antidepressant drugs have multiple therapeutic uses, such as treating mental and mood disorders including depression. However, research studies have indicated that a small segment of individuals, especially those below 25 years of age, may experience a worsening of their depressive symptoms or other mental/mood-related issues while taking antidepressants for any condition. There is also a possibility of experiencing suicidal thoughts or attempts in some cases.

3.2. Alternate Names of Drugs

Due to the existence of multiple brand names for a single drug across the world, social media data for a particular drug can vary depending on its usage location. Therefore, a list of drugs and their corresponding alternate names was created from Drugs.com [41] to ensure thorough coverage during the analysis. Table 3.1 shows the alternate names of the three drugs that were selected for this study.

Table 3.1 Alternate Names of Drugs

	Adderall	**Xanax**	**Prozac**
1.	Adderall XR	Alprazolam	Act Fluoxetine
2.	Aptensio XR	Clonazepam	Celexa
3.	Concerta	Diazepam	Citalopram
4.	Daytrana	Klonopin	Escitalopram
5.	Dexedrine	Niravam	Fluoxetine
6.	Dextrostat	Valium	Fluoxetine Hydrochloride
7.	Focalin	Xanax Ts	Fxt 10
8.	Metadate	Xanax XR	Lexapro
9.	Methylin		Paroxetine
10.	Mydayis		Paxil
11.	Ritalin		Pexeva
12.	Ritalin LA		Prozac Weekly
13.	Vyvanse		Rapiflux
14.			Sarafem

15.			Sertraline
16.			Zoloft

To gather a larger data set, the searches were performed with all alternative names of a single drug.

3.3. Misspelling of Drugs

One of the primary challenges encountered while using social media platforms is that users tend to misspell drug names. As shown in Figure 3.1, sample tweets containing the drug name Adderall were analyzed for correct and incorrect spellings. The correct spellings are highlighted in green, whereas the incorrect ones are shown in red. Therefore, relying solely on the correct spelling of a drug would limit the amount of data that can be obtained from social media networks. Consequently, a list of commonly misspelled drug names and those with phonetically similar spellings was compiled [42].

i take **aderall** for my adhd. i have 2 explain 2 them in blood tests since it has amphetamines in it. legal speed
zero sleep thanks to **adderall**
i am with you though, if nfl has amphetamines associated with **adderal** and vyvanse, fighters should not be allowed to do coke haha
not only antidepressants but other medications as well. i was misdiagnosed with adhd & was put on **aderol** and seroquel for 10 yrs
and if thats not bad enough she accomplished this with excessive amounts of **adderol** and vyvanse
which do you think currently has a higher demand, amphetamines (**aderoll**, vyvance) or cocaine? especially considering time of yr

Figure 3.1 Sample Tweets Containing Correct Spellings and Misspellings for the Drug Adderall

3.4. Dataset Collection

The data was collected using the multiple keywords of a drug. The data from Twitter was collected using the Tweepy and Twitterscraper APIs [43] which is provided by Twitter to aid the developers and researchers. The dataset contains 149,568 tweets that were filtered on three drugs with their generic names and alternate brand names. These prescription psychiatric medications include Adderall, Xanax, and Prozac, and deal with treatment of mental health issues. Tweets were gathered from July 2021 to August 2022. Figure 3.2 shows the distribution of tweets among the three classes.

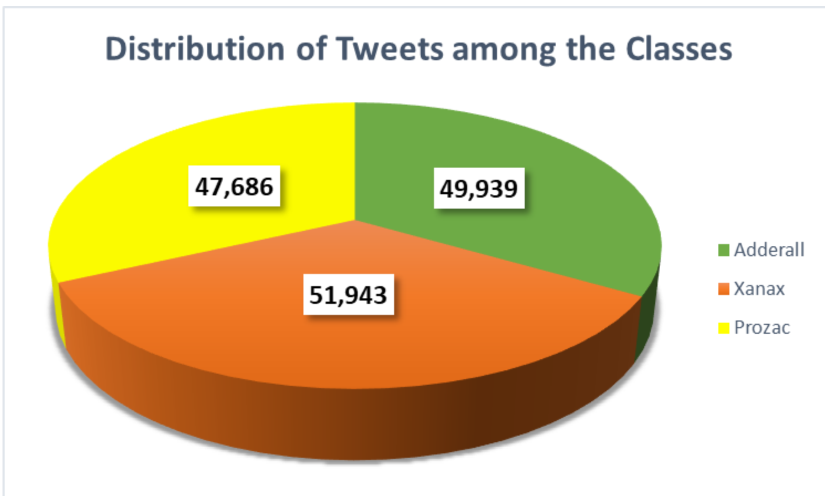

Figure 3.2 Distribution of Tweets among the Classes

3.5. Tools and Languages

Jupyter Notebook has been used as tool in this research for implementing the proposed methodology. It is an open-source web application that allows you to create and share interactive documents that contain live code, equations, visualizations, and narrative text. It is widely used in data science and scientific computing for exploratory data analysis, data visualization, machine learning, and scientific simulations.

Jupyter Notebook is a software application that runs on top of an operating system, and it does not directly provide RAM and disk space. So the RAM and disk space of the laptop were used. Table 5.1 in Chapter 5 shows the specifications of the laptop used.

Jupyter also supports Python that is used as a scripting language in this research. Python NLP and machine learning packages were imported into the Jupyter notebook by using package manager pip, which is included with Python. Just a single line of code was executed instead of downloading and installing them separately thereby saving a lot of time. Some of the packages used include the following:

- Keras with TensorFlow backend
- Numpy and pandas
- Matplotlib – For plotting
- csv – For reading csv files containing labels and metadata for the training set
- Importing NLP packages including nltk, SpaCy, WordCloud, STOPWORDS
- langdetect (detect & DetectorFactory)
- RegEx – For pattern matching and text manipulation
- Cleantext – For cleaning and normalizing text data
- defaultdict from collections – For detecting word frequency
- Genism (genism.model.phrases – For automatically detect common phrases)
- word2vec – For learning vector representations of words based on their co-occurrence in a corpus of text

Chapter 4: Methodology

This chapter provides a step-by-step account of the methodology of this study. Any challenges or limitations that were encountered along the way are also discussed in this chapter, and how they were addressed. Overall, this chapter aims to provide a clear and comprehensive overview of the methods and techniques used to design the proposed model. An overview of the process is provided in Figure 4.1.

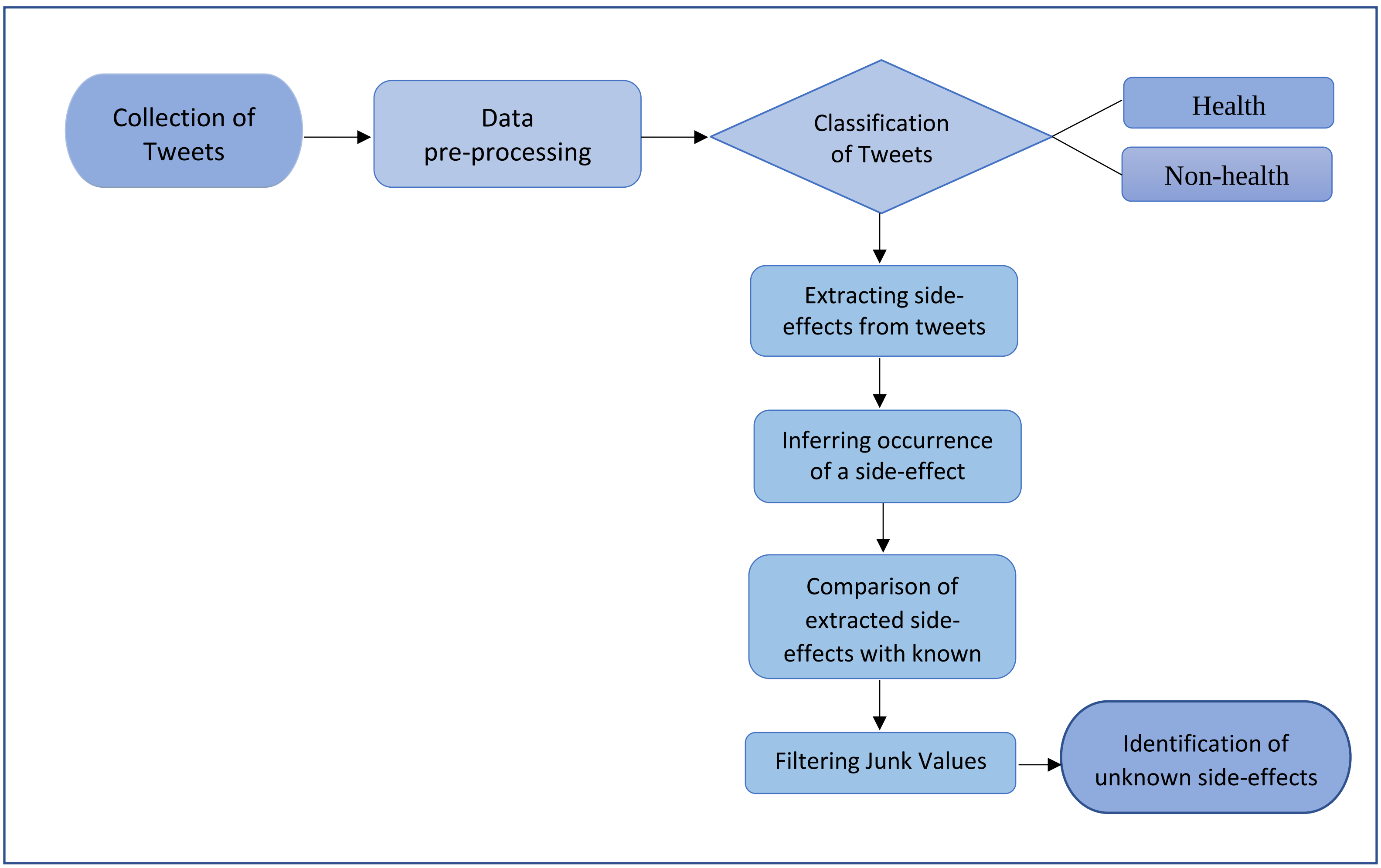

Figure 4.1 Overview of Proposed Methodology

4.1. Proposed Framework

Figure 4.1 shows the proposed framework. The framework consisted of eight steps. The first step was to capture the tweets. For this purpose, three drugs were shortlisted, their alternative names were gathered, and a list of common misspellings of the drugs and misspellings which were phonetically same as the drug were extracted. These lists were then used to collect a dataset of 149,568 tweets. This step has already been explained in detail in Chapter 3. The next step was to preprocess and clean the data. Three types of noise were removed from the data; tweets in foreign languages, retweets and redundant records followed by stemming and lemmatization. In the third step, the tweets were classified into "Health" and "Non-Health" tweets.

The fourth step was to extract side effects against the drugs from the tweets. For this purpose, word embeddings were generated using Word2Vec model. For the next step, it was needed to infer the value of occurrence of a particular side effect with a drug. The Cosine Similarity score was calculated and used. The next step was to compare the side effects extracted in this study with the known ones. A list of known side effects was taken from SIDER and WebMD. The junk values were then filtered from the extracted side effects and finally the unknown side effects of drugs were identified. Details of the proposed framework from the second step onwards are explained in this chapter. The detailed flow of the proposed methodology is shown in Figure 4.2.

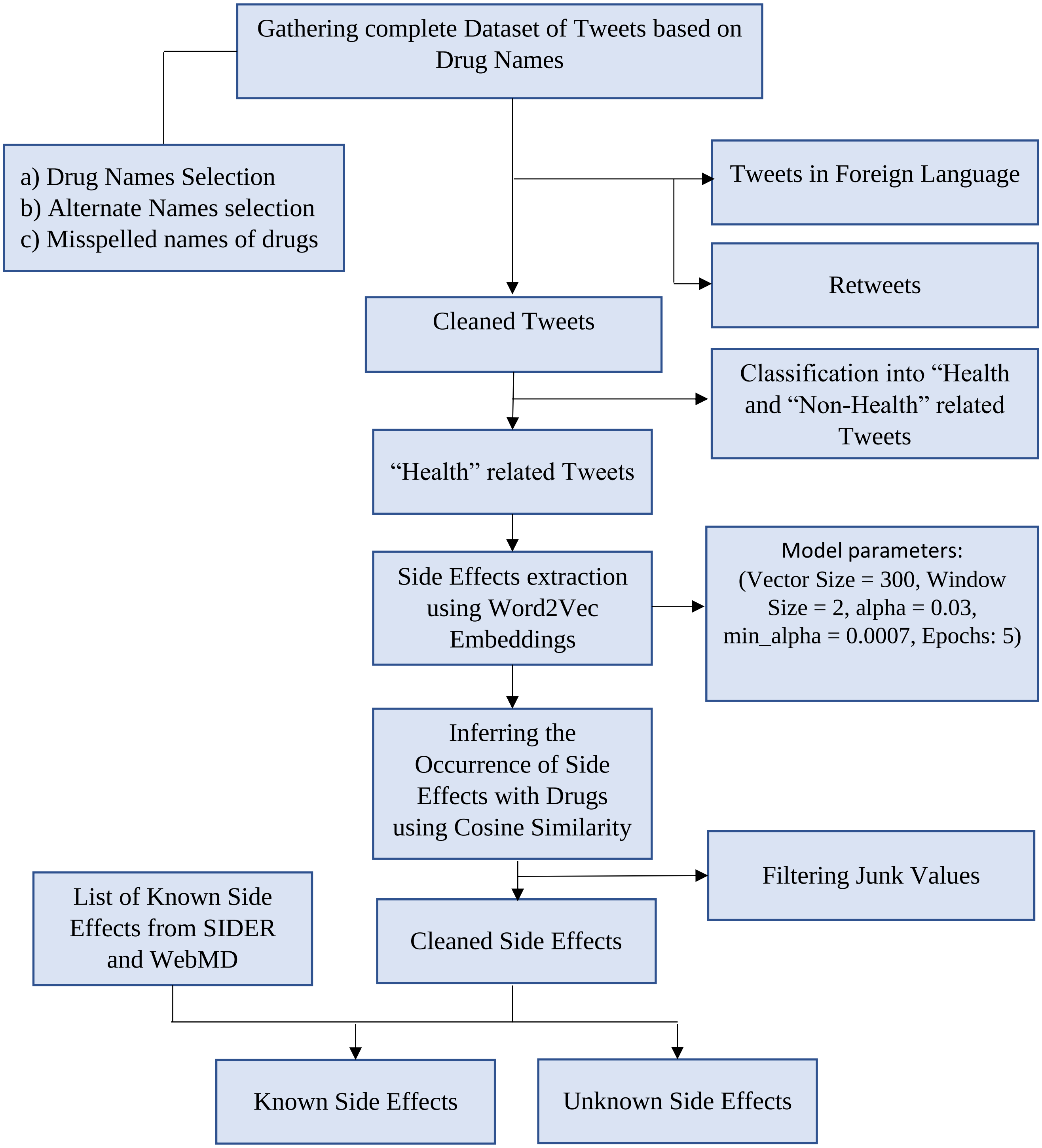

Figure 4.2 Detailed Flow of the Proposed Methodology

4.2. Data Preprocessing

The data obtained from Twitter comprised of a lot of noise. Three types of preprocessing was performed on the data to clean it:

- Removing Tweets in Foreign Languages
- Removing Retweets

Figure 4.3 and 4.4, show tweets in foreign languages and retweets respectively. This noise was removed using "langdetect" library. This library is a direct port of Google's language detection library and is used to filter out text written in non-target language. Duplicates are removed using the same.

me tiene tomando del Xanax
@persephonaae Lo xanax ??
RT @fechaessa: batem sempre na mm tecla, Ã© que oq ele tomou nÃ£o era xanax mas ya https://t.co/NUkkjZANln
Go listen toÂ… y me tiene tomando xanax solo por que no me ama
Me tiene tomando del xanax por dijo que no me ama
Xanax che goal avere una madre così.Xanax che goal avere una madre così.

Figure 4.3 Tweets in Foreign Languages

RT @bbyfagg: badthenny frankel threw the first xanax at
RT @CKnSD619: @BrookeBCNN I have PTSD, depression, panic disorder, and anxiety disorder. I can admit 10 years ago I once opened a bottle of...
RT @trishapaytas: ...and she is back !!! Lolol I fucking miss you kitty cat https://t.co/vLuoKCZC10 @gori_ognem ????? Xanax
RT @CKnSD619: @BeansPorkn I agree with you. At least we can admit it. The NRA would tells us no we need gunsto protect ourselves. If I had...
RT @luvmyselenur: 12. Quebonafide ft. PlanBe - Kawa i Xanax https://t.co/xBzXUqIjMX

Figure 4.4 Retweets

There were some junk tweets in the data as well. The tweets that contained drug names but no side effects were considered junk tweets for this study. Junk tweets were also removed.

After filtering out the noise and junk tweets, a total of 136,531 tweets were left. Figure 4.5 shows the distribution of cleaned tweets among the three classes.

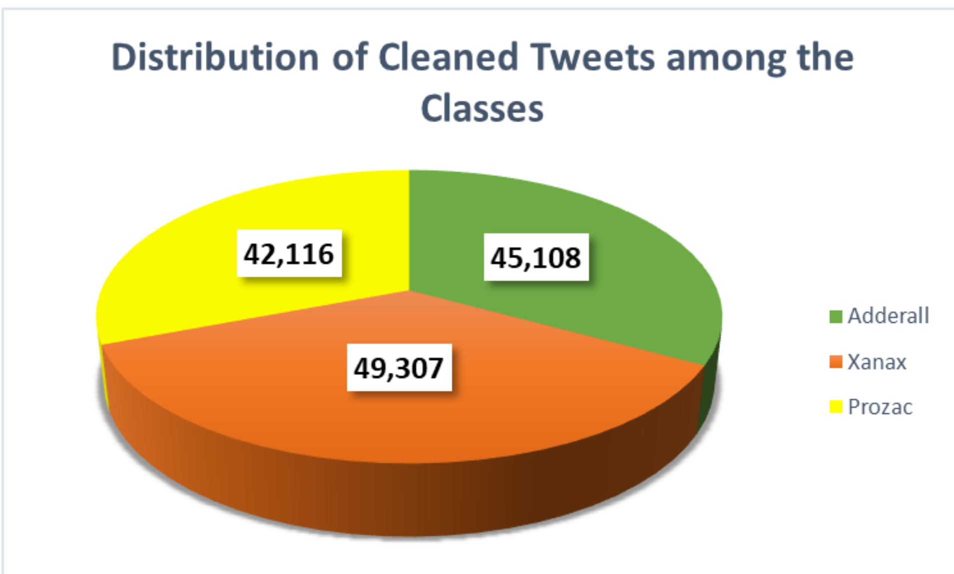

Figure 4.5 Distribution of Cleaned Tweets among the Classes

4.3. Classifying Health vs Non-Health Tweets

Given the unstructured nature of Twitter data, it's crucial to understand the context of each tweet. For instance:

i. "I took Adderall and it made it hard for me to sleep last night."

ii. "Does Adderall cause insomnia?"

The first tweet was classified as "Health" since it indicates the user experienced difficulty in sleeping after taking Adderall. In contrast, the second tweet is a question. It's possible for tweets to mention a drug name and its side effects in different contexts, so such tweets were categorized as "Non-Health". From the collected tweets 2500 random health related tweets were manually annotated as “Health” and 2500 were manually annotated as non-health. For training the machine learning model on manually annotated data, we used a used tf-idf with word2vec model, trained over general statements and questions, in order to filter out questions, queries or interrogative statements.

To minimize false positives, the tweets that fell under the "Non-Health" category were removed from the dataset for this study. After filtering out the “Non-Health” tweets, a total of 130,799 tweets were left. Figure 4.6 shows the distribution of “Health” related tweets among the three classes.

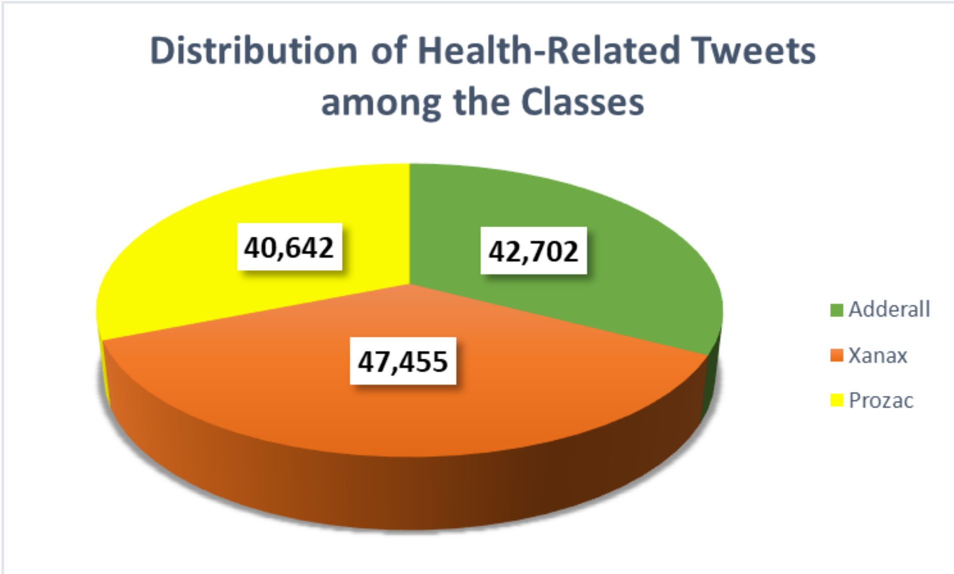

Figure 4.6 Distribution of Health-Related Tweets among the Classes

4.4. Extracting Side Effects

The next step was to extract side effects against each drug. Word embeddings were generated for this purpose using Word2Vec model. Word embeddings are a type of NLP technique used to represent words in a numerical format. Word embeddings map words to a high-dimensional vector space, where words with similar meanings are located closer to each other. This allows us to analyze and compare words based on their semantic meaning, rather than just their spelling or syntax.

Neural networks or other machine learning algorithms are trained to generate word embeddings by predicting the patterns and co-occurrence of words. Unstructured data and unsupervised learning techniques are used for this purpose. They generate a low dimensional and dense representation of words, capturing their contextual meaning and semantic relationships. For this study, the neural network model used for generating word embeddings was the Word2Vec model.

A group of Google researchers developed this model in 2013. Word2Vec model predicts the likelihood of words occurring in a specific scenario.

Semantic relationships and contextual meanings are used by Word2Vec to create high-dimensional vector representations of words. There are two methods for training a Word2Vec model; Continuous Bag-Of-Words (CBOW) and skip-gram. CBOW uses surrounding context words to predict the target word, whereas skip-gram uses the target word to predict context words. In this study, the CBOW model was utilized. The semantic relationships are stored used operations like addition and subtraction (e.g., "insomnia" - "antidepressant" + "sedative" = "drowsiness"). Words appearing in similar context are mapped to similar vectors.

There are different parameters that we need to set for the Word2Vec model. These include the size of vector i.e. the length of the feature vectors that represent each word in the model (generally 50-300), the window size i.e. the maximum distance that can be between the current and the predicted word in a sentence (generally 2-10), the minimum count i.e. the minimum frequency threshold for words to be included in the vocabulary (ignores all words with total absolute frequency lower than this, usually 2 to 10). Other parameters include alpha (initial learning rate), min_alpha (learning rate linearly drops to min_alpha as training progresses), negative (If > 0, negative sampling will be used, the int for negative specifies how many "noise words" should be drown. If set to 0, no negative sampling is used) and sample (The threshold for configuring which higher-frequency words are randomly downsampled. – usually between 0 to 1e-5). The parameters tuned for Word2Vec model for this study are mentioned in Table 4.1.

Table 4.1 Parameters Set for Word2Vec Model

Parameters	**Values**

Size of Vector	300
Window Size	2
Alpha (initial learning rate)	0.03
Min_alpha	0.0007
Negative	20
Sample	$6e^{-5}$
Minimum Count	20
Epochs	5

4.5. Inferring the Occurrence of a Side Effect in a Drug

The next step involved determining the frequency of occurrence of a specific side effect with a particular drug. To achieve this, the Cosine Similarity score was calculated for each drug with each side effect. The similarity between vectors in a high-dimensional space is measured using this score. It is a very common NLP technique for comparing similarity between sentences, paragraphs, and documents.

The Cosine Similarity score indicates the cosine of the angle between two vectors, which can range from -1 (completely dissimilar) to 1 (completely similar). A score of 0 indicates that the two vectors are orthogonal and have no similarity.

Cosine Similarity is calculated by normalizing two vectors to unit vectors, ensuring that their magnitudes are equal to 1. The dot product of the vectors is then determined by taking the sum of the element-wise products of the two vectors. The resulting dot product is divided by the product of the magnitudes of the two vectors, which yields the cosine similarity score. Mathematically, Cosine Similarity can be expressed as Eq. 4.1.

$$Cosine\ Similarity = Dot\ Product\ (V1, V2) / \ (Magnitude\ (V1) * Magnitude\ (V2)) \quad 4.1$$

where V1 and V2 are vector 1 and vector 2 respectively.

4.6. Extracting List of Known Side Effects

As the next step, a list of known side effects was extracted from the SIDER database [44] and WebMD [45]. This list contained all the known side effects of the three drugs that have been selected for this study i.e. Adderall, Xanax, and Prozac. A comprehensive list of these side effects is given in Table 4.2.

4.7. Filtering Junk Values from Extracted Side Effects

There were some junk values in the list of side effects extracted from the social media data. These junk values were filtered using a large phrasal ADR lexicon from FAERS (containing 20,285 phrases, complied by Hammad et. al [54]), where phrases representing the same ADRs were clustered.

4.8. Identification of Unknown Side Effects of Drugs

Finally, the side effects extracted from the Twitter dataset were compared with the list of side effects collected from the SIDER database and WebMD. The side effects that were in the former list but were missing in the latter one were the required unknown side effects of each drugs.

Table 4.2 List of Side Effects from the SIDER Database and WebMD

Sr. No	Adderall		Xanax		Prozac	
	SIDER Database	**WebMD**	**SIDER Database**	**WebMD**	**SIDER Database**	**WebMD**
1.	Abdominal cramps	Blood flow problems	Abdominal distress	Change in sex drive/ ability	Abdominal pain	Anxiety
2.	Abdominal pain	Blurred vision	Abnormal involuntary movements	Dizziness	Abnormal dreams	Black stools
3.	Acidosis	Change in sexual ability/ desire	Anxiety	Drowsiness	Abnormal ejaculation	Changes in sexual ability
4.	Aggression	Chest/ jaw/ left arm pain	Chest pain	Increased saliva pro duction	Amnesia	Decreased interest in sex
5.	Alopecia	Confusion	Cognitive disorder	Lightheadedness	Anorexia	Dizziness
6.	Anxiety	Continuous chewing	Confusion	Loss of coordination	Anxiety	Drowsiness

		movements/ teeth grinding				
7.	Body temperature increased	Diarrhea	Constipation	Memory problems	Asthenia	Easy bleeding/ bruising
8.	Breast disorder	Dizziness	Coordination abnormal	Mental/ mood changes	Depression	Eye pain/ swelling/ redness
9.	Cardiac death	Dry mouth	Decreased appetite	Trouble speaking	Dizziness	Fast heartbeat
10.	Chest pain	Extreme tiredness	Depression	Trouble walking	Dry mouth	Loss of appetite
11.	Chills	Fainting	Diarrhea		Dyspepsia	Mental/ mood changes
12.	Coma	Fast/pounding/irregular heartbeat	Drowsiness		Fatigue	Muscle weakness/ spasm
13.	Confusion	Fever	Dry mouth		Headache	Nausea
14.	Connective tissue disorder	Frequent/prolonged erections	Dysarthria		Increased appetite	Prolonged erection
15.	Constipation	Headache	Headache		Infection	Seizures

16.	Dry mouth	Loss of appetite	Hyperventilation		Influenza	Serious allergic reaction
17.	Eye disorder	Mental/ mood/ behavior changes	Increased appetite		Menstrual disorder	Shakiness (tremor)
18.	Fatigue	Nausea/ vomiting	Irritability		Mood swings	Signs of kidney problems
19.	Flushing	Nervousness	Libido decreased		Muscle twitching	Sweating
20.	Gastrointestinal disorder	Outbursts of words/ sounds	Lightheadedness		Nausea	Tiredness
21.	Hallucination	Rise in blood pressure	Memory impairment		Nervousness	Trouble sleeping
22.	Headache	Seizures	Menstrual disorder		Pain	Twitching muscles
23.	Infection	Shortness of breath	Rash		Pharyngitis	Unexplained fever
24.	Mental disorder	Stomach upset/ pain	Salivation		Rhinitis	Unusual agitation/ restlessness

25.	Rash	Swelling of the ankles/ feet	Somnolence		Sleep disorder	Unusual weight loss
26.	Sweating	Trouble sleeping	Strangury		Somnolence	Vision changes
27.	Vomiting	Trouble speaking	Sweating		Sweating	Vomit that looks like coffee grounds
28.	Weight decreased	Uncontrolled movements			Thinking abnormal	Widened pupils
29.		Unusual wounds			Tremor	Yawning
30.		Weakness on one side of the body			Upper respiratory tract infection	
31.		Weight loss			Urinary tract infection	
32.					Yawning	

Chapter 5: Experimentation & Results

This chapter discusses the experimental setup and results. The results are presented in various forms to analyze. Unknown as well as known side effects identified in this study have been discussed. A comparison of side effects with other studies is also given.

5.1. Experimental Setup

Jupyter Notebook has been used as tool in this research for implementing the proposed methodology. It is an open-source web application that allows you to create and share interactive documents that contain live code, equations, visualizations, and narrative text. It is widely used in data science and scientific computing for exploratory data analysis, data visualization, machine learning, and scientific simulations.

Jupyter Notebook is a software application that runs on top of an operating system, and it does not directly provide RAM and disk space. So the RAM and disk space of the laptop were used. Table 5.1 shows the specifications of the laptop used.

Table 5.1 Specifications of The Laptop Used

Processor	RAM	Disk Space	Operating System
Intel(R) Core(TM) i5-1035G1 CPU @ 1.00GHz 1.19 GHz	16.00 GB DDR4	512 GB SSD	Windows 10 Home

Jupyter also supports Python that is used as a scripting language in this research. Python NLP and machine learning packages were imported into the Jupyter notebook by using package manager pip, which is included with Python. Just a single line of code was executed instead of downloading and installing them separately thereby saving a lot of time.

5.2. Top Twenty Side Effects Identified for Each Drug

In the proposed methodology, side effects were extracted from tweets using Word2Vec word embeddings. The frequency of occurrence of a specific side effect with a particular drug was then determined. To achieve this, the Cosine Similarity score was calculated for each drug with each side effect. Using this Cosine Similarity score, a list of top twenty side effects identified for each drug respectively was formed. Figure 5.1, 5.2, and 5.3 show charts of top twenty side effects identified for Adderall, Xanax, and Prozac respectively along with their Cosine Similarity scores. Sinus, Dyslexia, and Dementia were the most common side effects of Adderall. For Xanax, the most common side effects were Insomnia, Panic, and Nausea. On the other hand, using Prozac caused GAD (Generalized Anxiety Disorder), OCD, and PMDD (Premenstrual dysphoric disorder) as the most common side effects. It is interesting to note here that Prozac is used to treat OCD and PMDD, but in some cases where the patient using Prozac did not have these conditions, using this drug induced these conditions in him according to the data collected from Twitter.

The results gathered through social media proved it to be a good medium to minimize the efforts of manual annotation and clinical trials of drugs. Clinical trials of drug is an expensive method which had its own restrictions on patient groups and drugs' usage. Manual annotation of the data is a tedious work to perform and hence takes good effort and time. Social media can help in reducing the cost of clinical trials and can help in gathering unknown side effects of drugs as well along with their known ones.

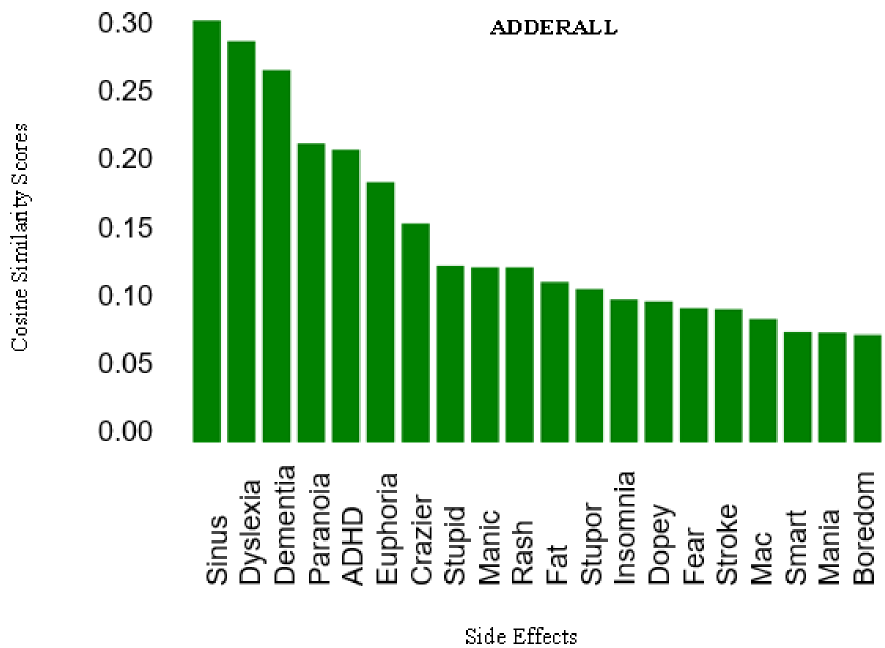

Figure 5.1 Top Twenty Side Effects Identified for Adderall along with Their Cosine Similarity Scores

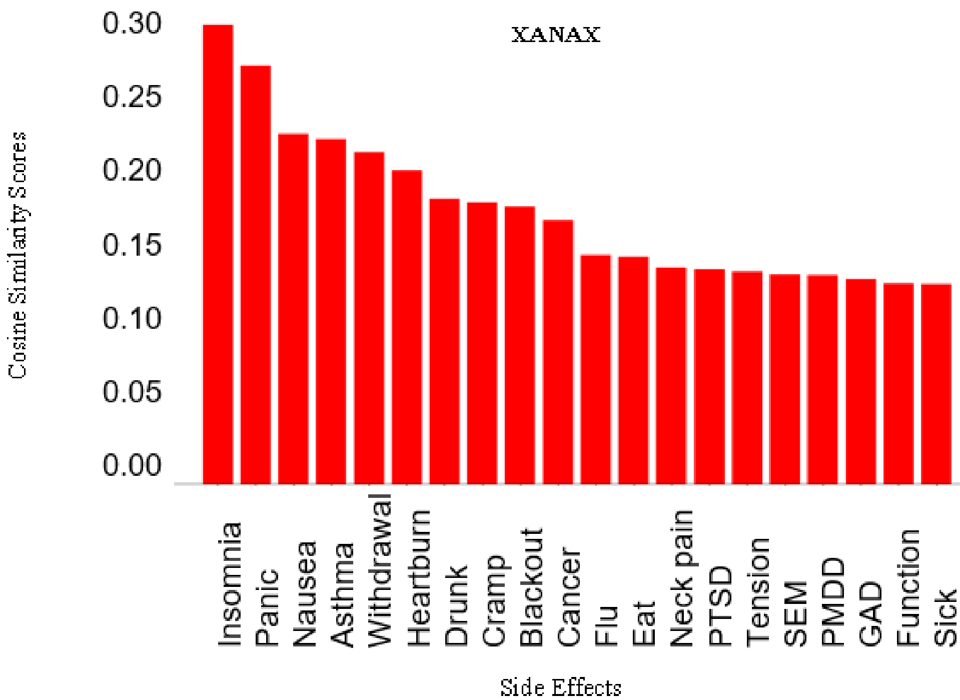

Figure 5.2 Top Twenty Side Effects Identified for Xanax along with Their Cosine Similarity Scores

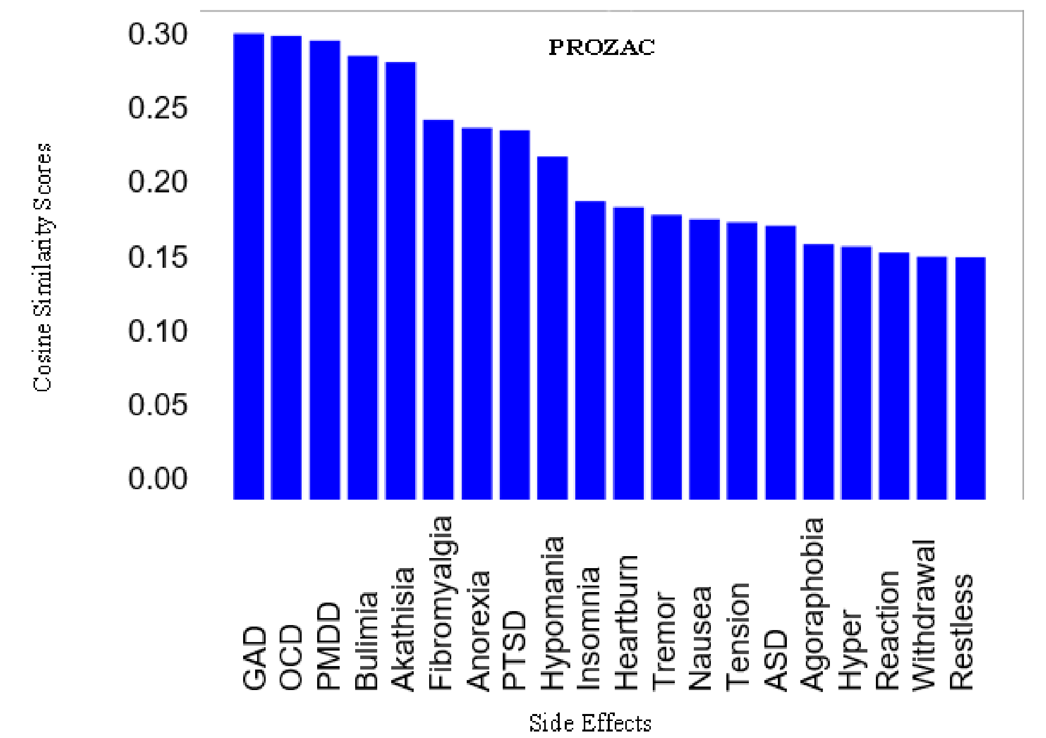

Figure 5.3 Top Twenty Side Effects Identified for Prozac along with Their Cosine Similarity Scores

5.3. Unknown Side Effects Identified in this Study

After conducting an analysis of the top twenty side effects associated with Adderall, Xanax, and Prozac, it was found that Adderall and Xanax had 1 unknown and 19 known side effects each, whereas Prozac had 3 unknown and 14 known side effects. The statistical model revealed some important findings that were not mentioned in the drug labels or databases for these drugs.

Specifically, it was found that Adderall was associated with **Dementia** (memory loss), Xanax was associated with **heartburn**, and Prozac was associated with **ASD** (Autism Spectrum Disorder), **Anorexia** (eating disorder that may lead to upset stomach), and **Bulimia** (eating disorder with uncontrolled episodes of over eating); patients may suffer from nightmares). These side effects were reported by users who had used the drugs but were not listed in the official drug labels or databases.

It is important to note that these findings are based on the analysis of this study and should be interpreted with caution. Further research is needed to confirm the existence and prevalence of these side effects.

5.4. Comparison of Side Effects with Previous Studies

According to Sarker *et al.* [46], Adderall is commonly abused for diarrhea induced weight loss via social media. This study found similar results, with diarrhea having similarity value of 0.049 with Adderall. Smith *et al.* [47] compared ADRs reported on social media with those in traditional sources, finding that stroke-related headache was commonly reported in FAERS and drug information databases. This study also found strokes mentioned on Twitter, with similarity value of 0.093. In Chavant *et al.*'s [48] French PharmacoVigilance Database, memory disorders were reported for Xanax and Prozac at rates of 14 and 16, respectively. Our methodology also found high Cosine Similarity score for memory disorders as an ADR for Xanax and Prozac. Conditions related to Xanax include PTSD (score = 0.098), PMDD (score = 0.095), Insomnia (score = 0.020), all of which are contributing medical conditions towards memory loss. Conditions related to Prozac include GAD (score = 0.093), OCD (score = 0.32), PMDD (score = 0.36) and PTSD (0.34), all of which are contribute to memory loss.

The potential link between Prozac and aggressive behavior has been a subject of debate over the last thirty years [49]. However, the analysis of data from Twitter done in this study suggests that some patients may experience increased aggression after taking Prozac with hyper behavior having a score of 0.37 and hypomania (associated with physical aggression) having a score of (0.12).

Another commonly reported side effect of Prozac, as shown by the analysis done in this study is the occurrence of withdrawal syndrome (score = 0.12). This side effect has been previously documented in various studies [52, 53]. Additionally, bulimia (digestive problems and upset stomach issues as a result of binge eating disorder) was reported in this study with score = 0.061, and has also been found in a study involving preschool and high school children [52].

Heartburn was predicted as an under-reported side effect of Xanax with a score of 0.14. This is supported by a previous clinical trial where more than 30% of the patients reported experiencing heartburn [53]. Furthermore, this study's findings on Adderall that it induced memory loss (Dementia as a side-effect with score = 0.26) are supported by a previous study [55]. Table 5.2 shows a summary of the similarity scores of the common ADRs reported in this study and the previous studies.

Table 5.2 Summary of the Similarity Scores of the Common ADRs Reported in this Study and the Previous Studies

ADRs	**Drugs**	**Similarity Score in This Study**	**Side-effects found in Previous Studies**	
			Studies	**Side effects**
Weight Loss	Adderall	0.049	Sarker *et al.* [46]	Weight Loss
Strokes	Adderall	0.093	Smith *et al.* [47]	Stroke-related headache
Memory Disorder (PTSD)	Xanax	0.098	Chavant *et al.* [48]	Memory disorder
	Prozac	0.34		Memory disorder
Memory	Xanax	0.098	Chavant *et al.* [48]	Memory disorder

Disorder (PMDD)	Prozac	0.36		Memory disorder
Obsessive Compulsive Disorder (OCD)	Prozac	0.32	Chavant *et al.* [48]	Memory Disorder
Generalized Personality disorder (GAD)	Prozac	0.093	Chavant *et al.* [48]	Memory Disorder
Hyper behavior	Prozac	0.37	O'Connor *et al.* [49]	Aggression
Withdrawal Syndrome	Prozac	0.12	Barterian *et al.* [52]	Withdrawal Syndrome
			Singh *et al.* [53]	Withdrawal syndrome
digestive and stomach issues	Prozac	0.061	Barterian *et al.* [52]	Stomach issues
Heartburn	Xanax	0.14	Unreported	
Dementia	Adderall	0.26	Unreported	
Builimia	Prozac	0.061	Unreported	
Anorexia	Prozac	0.010	Unreported	
ASD	Prozac	0.44	Unreported	

Chapter 6: Conclusion & Future Work

Sometimes, when a patient is given medication, it can cause some harmful and undesirable side effects. These are called ADRs. ADRs are a significant public health concern around the world. In order to reduce or alleviate the risk of developing ADRs, healthcare professionals take necessary steps to identify, manage, and prevent them. ADRs are underreported through conventional reporting methods. With the increasing popularity of social media platforms, it has become a hub of information. It is observed that ADRs are also being discussed there. Many experiences of patients and doctors with medicines can be found on social media posts.

This study proposed an approach to detect and list unknown ADRs from social media data using Word2Vec and Cosine Similarity techniques. The framework utilized Natural Language Processing (NLP) to automate the discovery of ADRs mentioned in social media posts. They were then compared to a list of known ADRs to identify unknown ADRs. The dataset for this study had been self-collected and contained tweets related to ADRs. Three drugs were shortlisted for this study; Adderall, Xanax, and Prozac. For Adderall and Xanax, one unknown ADR each was found; Dementia (score = 0.26) and heartburn (score = 0.14) respectively, whereas, for Prozac, three ADRs with relatively high similarity values were found; ASD (0.044), Anorexia (0.010), and Bulimia (0.061).

While this study provides a valuable contribution to the field, there is some potential future work for this study. In the future, some newer techniques like BERT and transformer-based models can be used. Secondly, for this study, the data was collected from Twitter only. In the future, the data can be expanded by including data from other social media platforms as well, like Facebook and Reddit. Moreover, the demographics were not included as part of this study. Incorporating them along with the social media posts can help to investigate the variation of ADRs in different populations and regions in the future. lastly, the scope of this study was limited to three drugs; Adderall, Xanax, and Prozac. In the future, the work can be expanded for other drugs also.

References

[1] World Health Organization, “Safety of Medicines: A Guide to Detecting and Reporting Adverse Drug Reactions: Why Health Professionals Need to Take Action,” 2022.

[2] E. Al-Shareef *et al.*, “Detection of Adverse Drug Reactions in COVID-19 Hospitalized Patients in Saudi Arabia: A Retrospective Study by ADR Prompt Indicators,” *Healthc.*, vol. 11, no. 5, 2023, doi: 10.3390/healthcare11050660.

[3] Turing, “How Does Natural Language Processing Function in AI?,” 2022. https://www.turing.com/kb/natural-language-processing-function-in-ai.

[4] B. C. Wallace and Y. Zhang, “A Sensitivity Analysis of (and Practitioners’ Guide to) Convolutional Neural Networks for Sentence Classification,” *arXiv Prepr. arXiv1510.03820*, 2015.

[5] Y. Shen, X. He, J. Gao, L. Deng, and G. Mesnil, “Learning Semantic Representations Using Convolutional Neural Networks for Web Search,” in *Proceedings of the 23rd international conference on world wide web*, 2014, pp. 373–374.

[6] A. Mohammed and R. Kora, “An Effective Ensemble Deep Learning Framework for Text Classification,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 10, pp. 8825–8837, 2022, doi: 10.1016/j.jksuci.2021.11.001.

[7] Y. Liu, P. Li, and X. Hu, “Combining Context-Relevant Features with Multi-Stage Attention Network for Short Text Classification,” *Comput. Speech Lang.*, vol. 71, p. 101268, 2022, doi: 10.1016/j.csl.2021.101268.

[8] C. Grouin *et al.*, “Descriptions of Adverse Drug Reactions Are Less Informative in Forums

Than in the French Pharmacovigilance Database but Provide More Unexpected Reactions," *Front. Pharmacol.*, vol. 9, pp. 1–11, 2018, doi: 10.3389/fphar.2018.00439.

[9] Y. Zhang *et al.*, "Applying a deep learning-based sequence labeling approach to detect attributes of medical concepts in clinical text," *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 5, pp. 1–8, 2019, doi: 10.1186/s12911-019-0937-2.

[10] M. Gaur, S. Scepanovic, E. Martin-Lopez, D. Quercia, and K. Baykaner, "Extracting Medical Entities from Social Media," in *Proceedings of the ACM Conference on Health, Inference, and Learning*, 2020, pp. 170–181, doi: 10.1145/3368555.3384467.

[11] M. Tiftikci, A. Özgür, Y. He, and J. Hur, "Machine learning-based identification and rule-based normalization of adverse drug reactions in drug labels," *BMC Bioinformatics*, vol. 20, no. 21, pp. 1–9, 2019, doi: 10.1186/s12859-019-3195-5.

[12] Y. Zhang, Q. Chen, Z. Yang, H. Lin, Z. Lu, and B. Wallace, "Adverse drug event detection via deep learning-based natural language processing," *Comput. Biol. Med.*, vol. 99, pp. 157–164, 2018.

[13] Y. Li, H. Lin, Z. Yang, J. Wang, and J. Luo, "An NLP-based hybrid approach for detecting adverse drug reactions from clinical notes," *J. Biomed. Inform.*, vol. 94, no. 103193, 2019.

[14] H. Xu, Z. Lu, and J. Hao, "Graph neural networks for adverse drug event extraction from biomedical literature," *J. Biomed. Inform.*, vol. 105, no. 103406, 2020.

[15] D. Hazarika, A. D. Farmer, P. Gohil, J. Lee, and J. C. Denny, "An NLP-Based System for Adverse Drug Reaction Detection in Twitter Posts," *Stud. Health Technol. Inform.*, vol. 281, pp. 63–67, 2021.

[16] M. Liu, Y. Wu, Y. Chen, Y. Wang, and Y. Li, “A multi-task learning framework for adverse drug event detection in electronic health records,” *J. Biomed. Inform.*, vol. 116, no. 103690, 2021.

[17] D. S. Miranda, “Automated detection of adverse drug reactions in the biomedical literature using convolutional neural networks and biomedical word embeddings,” *arXiv Prepr. arXiv1804.09148*, 2018.

[18] Y. Wu, M. Jiang, and J. Lei, “Adverse drug event detection using deep learning with convolutional neural networks,” *J. Biomed. Inform.*, vol. 76, pp. 103–112, 2017.

[19] Y. Zhang, Q. Chen, Z. Yang, H. Lin, Z. Lu, and B. Wallace, “Adverse drug event detection via deep learning-based natural language processing,” *Comput. Biol. Med.*, vol. 99, pp. 157–164, 2018.

[20] H. X. Tan *et al.*, “Combining machine learning with a rule-based algorithm to detect and identify related entities of documented adverse drug reactions on hospital discharge summaries,” *Drug Saf.*, vol. 45, no. 8, pp. 853–862, 2022, doi: 10.1007/s40264-022-01196-x.

[21] X. Li, Q. Sun, S. Wang, Y. Huang, and S. Liu, “Learning semantic representations for adverse drug event detection using BERT,” *BMC Med. Inform. Decis. Mak.*, vol. 21, no. 1, pp. 1–13, 2021.

[22] A. J. Masino, D. Forsyth, and A. G. Fiks, “Detecting Adverse Drug Reactions on Twitter with Convolutional Neural Networks and Word Embedding Features,” *J. Healthc. Informatics Res.*, vol. 2, no. 1–2, pp. 25–43, 2018, doi: 10.1007/s41666-018-0018-9.

[23] Y. Gao, J. Chen, and Y. Chen, "An NLP-based method for detecting adverse drug reactions from clinical notes," *J. Biomed. Inform.*, vol. 113, no. 103625, 2021.

[24] Y. Zhang, S. Cui, and H. Gao, "Adverse drug reaction detection on social media with deep linguistic features," *J. Biomed. Inform.*, vol. 106, no. April, p. 103437, 2020, doi: 10.1016/j.jbi.2020.103437.

[25] A. Nikfarjam, A. Sarker, K. O'Connor, R. Ginn, and G. Gonzalez, "Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features," *J. Am. Med. Informatics Assoc.*, vol. 22, no. 3, pp. 671–681, 2015.

[26] T. Teklemariam, A. Nguyen, L. Martin, and S. Ji, "Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts," *J. Am. Med. Informatics Assoc.*, vol. 25, no. 9, pp. 1175–1181, 2018.

[27] J. Bian, U. Topaloglu, F. Yu, and J. Lane, "Towards large-scale twitter mining for drug-related adverse events," in *Proceedings of the 2015 IEEE International Conference on Healthcare Informatics*, 2015, pp. 460–469.

[28] R. Ginn *et al.*, "Mining Twitter for adverse drug reaction mentions: A corpus and classification benchmark," in *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, 2017, pp. 56–65.

[29] I. Ahmed, M. Ahmad, and M. Alghamdi, "A deep learning approach for adverse drug event detection from social media data," *IEEE Access*, vol. 6, pp. 60097–60105, 2018.

[30] J. Chen, L. He, and Y. Zhang, "Adverse drug event detection from Twitter data using hybrid

feature sets," *J. Biomed. Inform.*, vol. 91, no. 103120, 2019.

[31] P. Pimpalkhute, A. Nikfarjam, A. Patki, K. O'Connor, A. Sarker, and G. Gonzalez, "Empowering health care clients with conversational agents and social media mining," in *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, 2017, pp. 56–65.

[32] A. Yates, N. Goharian, and O. Frieder, "ADRTrace: detecting expected and unexpected adverse drug reactions from user reviews on social media sites," *J. Biomed. Inform.*, vol. 65, pp. 74–86, 2017.

[33] J. Li, H. Liu, J. Zhao, and Y. Chen, "Adverse drug reaction detection using deep learning methods and sources of medical information," *Healthc. Technol. Lett.*, vol. 7, no. 3, pp. 85–90, 2020.

[34] L. Li and J. Huang, "Adverse drug event detection using deep learning coupled with semantic embedding of unstructured clinical text," *IEEE J. Biomed. Heal. informatics*, vol. 23, no. 2, pp. 674–681, 2018.

[35] M. Yang, M. Kiang, W. Shang, and A. Filtering, "Preprocessing social media data for signal processing and classification applications: A survey," *Inf. Fusion*, vol. 14, no. 4, pp. 557–571, 2019.

[36] G. Lample and F. Yvon, "Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions," in *26th International Conference on Computational Linguistics: Technical Papers*, 2016, pp. 866–875.

[37] R. W. White, N. P. Tatonetti, N. H. Shah, R. B. Altman, and E. Horvitz, "Web-scale pharmacovigilance: listening to signals from the crowd," *J. Am. Med. Informatics Assoc.*,

vol. 20, no. 3, pp. 404–408, 2013.

[38] I. Korkontzelos, A. Nikfarjam, M. Shardlow, A. Sarker, S. Ananiadou, and G. Gonzalez, “Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts,” *J. Biomed. Inform.*, vol. 62, pp. 148–158, 2016.

[39] Y. Wang, C. Huang, Y. Peng, M. Guo, and X. Hu, “An ensemble deep learning framework for drug adverse event detection from social media data,” *J. Biomed. Inform.*, vol. 88, pp. 20–28, 2018.

[40] A. Mavragani, G. Ochoa, and K. P. Tsagarakis, “Assessing the methods, tools, and statistical approaches in Google Trends research: systematic review,” *J. Med. Internet Res.*, vol. 22, no. 11, p. 22184, 2020.

[41] “Prescription drug information,” *Drugs.com*. https://www.drugs.com/ (accessed Apr. 07, 2023).

[42] A. Sarker and G. Gonzalez, “A corpus for mining drug-related knowledge from Twitter chatter: Language models and their utilities,” *Data Br.*, vol. 10, pp. 122–131, 2017, doi: 10.1016/j.dib.2016.11.056.

[43] “Tweepy,” *Tweepy*. http://www.tweepy.org/ (accessed Apr. 09, 2023).

[44] Sider 4.1, “Sider SideEffects.” http://sideeffects.embl.de/se/ (accessed Apr. 11, 2023).

[45] WebMD, “WebMD - Better information. Better health.” https://www.webmd.com/ (accessed Apr. 11, 2023).

[46] A. Sarker *et al.*, “Social media mining for toxicovigilance: Automatic monitoring of prescription medication abuse from twitter,” *Drug Saf.*, vol. 39, no. 3, pp. 231–240, 2016,

doi: 10.1007/s40264-015-0379-4.

[47] K. Smith, S. Golder, A. Sarker, Y. Loke, K. O'Connor, and G. Gonzalez-Hernandez, "Methods to Compare Adverse Events in Twitter to FAERS, Drug Information Databases, and Systematic Reviews: Proof of Concept with Adalimumab," *Drug Saf.*, vol. 41, no. 12, pp. 1397–1410, 2018, doi: 10.1007/s40264-018-0707-6.

[48] F. Chavant, S. Favrelière, C. Lafay-Chebassier, C. Plazanet, and M. C. Pérault-Pochat, "Memory disorders associated with consumption of drugs: Updating through a case/noncase study in the French PharmacoVigilance Database," *Br. J. Clin. Pharmacol.*, vol. 72, no. 6, pp. 898–904, 2011, doi: 10.1111/j.1365-2125.2011.04009.x.

[49] K. O'Connor, P. Pimpalkhute, A. Nikfarjam, R. Ginn, K. L. Smith, and G. Gonzalez, "Pharmacovigilance on twitter? Mining tweets for adverse drug reactions," *AMIA Annu. Symp. Proc.*, vol. 2014, pp. 924–933, 2014.

[50] A. Wichniak, A. Wierzbicka, M. Walęcka, and W. Jernajczyk, "Effects of Antidepressants on Sleep," *Curr. Psychiatry Rep.*, vol. 19, no. 9, pp. 1–7, 2017, doi: 10.1007/s11920-017-0816-4.

[51] J. M. Parish, "Violent dreaming and antidepressant drugs: Or how paroxetine made me dream that i was fighting Saddam Hussein," *J. Clin. Sleep Med.*, vol. 3, no. 5, pp. 529–531, 2007, doi: 10.5664/jcsm.26919.

[52] J. A. Barterian, E. Rappuhn, E. L. Seif, G. Watson, H. Ham, and J. S. Carlson, "Current state of evidence for medication treatment of preschool internalizing disorders," *Sci. World J.*, vol. 2014, 2014, doi: 10.1155/2014/286085.

[53] S. Singh, R. T. Bailey, H. J. Stein, T. R. DeMeester, and J. E. Richter, “Effect of Alprazolam (Xanax) on Esophageal Motility and Acid Reflux,” *Am. J. Gastroenterol. (Springer Nature)*, vol. 87, no. 4, 1992.

[54] Farooq Hammad, Naveed Hammad. *2019 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)* IEEE; 2019. Gpadrlex: Grouped phrasal adverse drug reaction lexicon; pp. 1–6.

[55] L. Sanday, C. L. Patti, K. A. Zanin, S. Tufik, and R. Frussa-Filho, “Amphetamine-induced memory impairment in a discriminative avoidance task is state-dependent in mice,” *Int. J. Neuropsychopharmacol.*, vol. 16, no. 3, pp. 583–592, 2013, doi: 10.1017/S1461145712000296.