TRANSFER LEARNING BASED COMPUTATIONAL FRAMEWORK FOR ANALYSING USER REVIEWS TO DETECT ADVERSE DRUG REACTIONS USING BIO-BERT

Final Year Project Report

submitted in partial fulfilment of the requirements for the award of the degree

of

Bachelor of Technology

in

COMPUTER SCIENCE & ENGINEERING

by

Sushmita Paul (B18CS007)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
NATIONAL INSTITUTE OF TECHNOLOGY MEGHALAYA, INDIA

MAY, 2022

CERTIFICATE

I hereby certify that the work which is presented in the Final Year Project report titled "TRANSFER LEARNING BASED COMPUTATIONAL FRAMEWORK FOR ANALYSING USER REVIEWS TO DETECT ADVERSE DRUG REACTIONS USING BIO-BERT", in partial fulfillment of the requirements for the award of the **Bachelor of Technology in Computer Science & Engineering** and submitted to the Department of Computer Science & Engineering of National Institute of Technology Meghalaya, India is an authentic record of our own work carried out during a period from August, 2021 to May, 2022 under the supervision of **Dr. Yogita, Assistant Professor**.

The matter presented in this report has not been submitted by me for the award of any other degree elsewhere.

Sushmita Paul
B18CS007

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Date: 26 May, 2022

(Signature of Supervisor)
Dr. Yogita
Assistant Professor
NIT Meghalaya

Head
Department of Computer Science & Engineering,
National Institute of Technology Meghalaya, India

DECLARATION OF ORIGINALITY

I hereby declare that this project work titled "TRANSFER LEARNING BASED COMPUTATIONAL FRAMEWORK FOR ANALYSING USER REVIEWS TO DETECT ADVERSE DRUG REACTIONS USING BIO-BERT" represents my original work carried out as a student of the Department of Computer Science & Engineering of National Institute of Technology Meghalaya, India and to the best of our knowledge it contains no material previously published or written by another person unless cited. Any contribution made to this project work by others with whom we have worked at National Institute of Technology Meghalaya or elsewhere, is explicitly acknowledged.

Sushmita Paul

B18CS007

Place: Shillong, Meghalaya

Date: 26 May, 2022

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Dr. Yogita, Assistant Professor, Department of Computer Science and Engineering, NIT Meghalaya for the continuous support during my internship study and related research, for her patience, motivation, and immense knowledge. Besides her guidance at all the times of research and writing of this report, the insightful comments and questions stimulated me to widen my research from various perspectives. I could not have imagined having a better mentor for the project.

I also thank Mr. S. R. Ngamwal Anal, PhD scholar at NIT Meghalaya, for his valuable perceptive inputs during the project duration.

My sincere thanks also goes to Dr. Deepak Kumar and the Department of Computer Science and Engineering, NIT Meghalaya for giving us all technical support and cooperation online amidst pandemic and offline too.

I thank my friends at my institute and school for enlightening me and being a driving force the entire time.

Last but not the least, I would like to thank my family for supporting me spiritually, mentally and emotionally throughout.

Sushmita Paul

B18CS007

TRANSFER LEARNING BASED COMPUTATIONAL FRAMEWORK FOR ANALYSING USER REVIEWS TO DETECT ADVERSE DRUG REACTIONS USING BIO-BERT

Abstract

In the current state-of-the-art, there are numerous text mining techniques, but not much for pharmaceutical analysis or on biomedical data. Moreover, on drugs reviews provided by patients on online platforms like `www.webmd.com`, we may find some sentiment analysis articles, satisfaction analysis of drugs or opinion mining of the user reviews. In this paper, we analyse text drugs' reviews of six drugs, namely, Acetaminophen, Naproxen, Sumatriptan, Carbamazepine, Lamotrigine and Levetiracetam, where the former three are used for the treatment of migraine while the latter three are that of seizure, and their various types. We present a computational framework that has been built using transfer learning techniques on Bio-BERT pre-trained model. Bio-BERT is a model that is trained on several biomedical and english literature text. It evolves from the BERT model which was only trained on english literature texts. We further fine-tune our model to analyse the transfer learning capacities and efficiencies over cross-domain drugs, on different combinations of the dataset. Finally, we report the ADRs detected by our model, the various ADRs that are detected but not in the SIDER, analyse how and why severity is detected by our model in certain cases, and analyse the top 10 ADRs detected by our model for each drug as a proof of correctness task with respect to websites like `https://pubmed.ncbi.nlm.nih.gov/` and `https://medlineplus.gov/`.

List of Figures

List of Tables

Table of Contents

1 Introduction

Biomedical text mining is becoming increasingly important as the number of documentations on it rapidly grows. Extracting valuable information from biomedical literature has gained popularity among researchers, and deep learning has boosted the development of effective biomedical text mining models. While we know neuropathic dysfunctions do not usually leave any physical bruises, the side effects of drug treatment can be hazardous and may cost lifelong disability or dysfunctions. It is, therefore, very necessary to bring forward the adverse drug reactions (ADRs) of common medications prescribed and widely used across the globe. Adverse drug reactions, according to the definition of the World Health Organization (WHO), are a response to noxious medication which occurs as a result of normal doses used in man for diagnosing or curing a disease [4]. In simple words, ADRs are the side-effects of a specific drug. Side-effects are one of the top causes of morbidity and mortality.

1.1 Motivation

The very commonly used medicinal components of drugs are tested and trusted over the years. Thus, it only seems logical to find the ADRs of drugs that are between the top 5 to 10 listed drugs for a disorder according to domain experts.

Our objective is to focus primarily on a similar domain of disorders, namely, migraine and seizures. They may be segregated as episodic disorders, whereby patients are generally normal but they suffer from spontaneous attacks on varying timeline depending on the severity of the case. Both migraine and seizures involve abnormal electrical activity in the brain. In case of migraines, nerve cells are at very hyperpolarized levels while on the other hand, seizures occur due to hypersynchronous discharge of neurons. Migraine may be caused due to various factors ranging from hormonal disorder, emotional stress, physical, dietary, environmental to medicinal changes. These causes are different from individual to individual. Migraines may be of several types. Some of them are migraines with or without aura, chronic aura, migraine with brainstem aura, vestibular migraine, hemiplegic migraine, abdominal migraine, medicinal overuse migraine and so on.

Seizures are single occurrences of a neurological condition triggering due to electrical abnormality in nervous system. The repeated occurrence of seizures is a condition of epilepsy. Seizures may be of different categories like simple and complex partial focal seizures, tonic-clonic seizures, absence seizure, myoclonic seizure, clonic seizure, atonic seizure and status epilepticus. They may be majorly triggered due to improper sleep cycles, stress, abruptly waking up, or drinking alcohol.

On the recommendation of medical practitioners and survey on popular websites like `https://www.webmd.com/migraines-headaches/migraine-treatment` , `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4110862/` and `https://www.rxlist.com/seizure_medications/drugs-condition.htm` for ensuring the right drugs to be chosen for our study. We chose the following six drugs, three drugs each for the treatment of migraine and seizures:

- Acetaminophen
- Naproxen
- Sumatriptan
- Carbamazepine
- Lamotrigine
- Levetiracetam

2 Literature Survey

2.1 State-of-the-art

Over the years, many technologies have emerged for text mining. Through sentiment analysis using convolutional neural network (CNN) or natural language processing (NLP) or other technologies. In [6], a comparison based analysis of deep learning models are done and their importance in pharmaceutical discovery is explained. The analysis of various deep learning architectures like Long short-term memory (LSTM) recurrent neural network or CNN on sentiment analysis of drug reviews is stated in [1]. Some related work on cross-domain sentiment analysis of drug reviews over positiveness, negativeness and over effectiveness can be seen in [3]. Researchers have also exploited opinion mining techniques on texts available on online platforms or social media. However, these methods

are used to extract the overall satisfaction or general opinion of patients on drugs. The model in [2] states that it outperforms support vector machine (SVM) on opinion mining. Transfer learning on drug reviews on trusted sites that are available online openly for the public, for the detection of ADRs is a comparatively less sorted aspect of research.

2.2 Word Embedding

An embedding is a matrix on the size of vocabulary and the number of dimensions in the low dimensional space. These embeddings are normally trained from unlabelled text that are usually redundant in huge amounts from sources like Wikipedia. The embeddings are usually trained in a fashion so that the dot product of vectors of a word and its neighbour word preserves the words' point-wise mutual information. After being trained, these vectors can be used to look for word synonyms by looking for words with their vectors closest to the searched word's vector. By representing words using these vectors, the model captures derived information from co-occurrences of the contained words from the unsupervised pre-training. Additionally using lower dimensional vector space also helps reduce overfitting. A tokenised sentence with their tokens projected by an embedding becomes a dense matrix that can then be fed as an input into a neural network [5].

2.3 BioBERT

BERT (Bidirectional Encoder Representations from Transformers) involves two tasks, pre-training and fine-tuning [4]. BioBERT (BERT for Biomedical Text Mining) is a large-scale biomedical corpus-trained domain-specific language representation model. BioBERT beats BERT and earlier state-of-the-art models in a range of biomedical text mining tasks when pre-trained on biomedical corpora, thanks to its nearly same architecture across tasks. BioBERT beats prior state-of-the-art models on three sample biomedical text mining tasks: biomedical named entity recognition, biomedical connection extraction, and biomedical question answering, whereas BERT achieves performance equivalent to previous state-of-the-art models [7]. BioBERT is pre-trained on PubMed abstracts and PMC full-text articles in addition to the general English Wikipedia and BooksCorpus, and further fine-tuned over the biomedical text mining tasks. In our project, we have the *named entity recognition technique to find ADRs* from user reviews of drugs given by

patients.

2.4 Transfer Learning

Transfer learning (TL) is a research problem in machine learning that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem. For example, knowledge gained while learning to recognize cars could apply when trying to recognize trucks. This area of research bears some relation to the long history of psychological literature on transfer of learning, although practical ties between the two fields are limited. From the practical standpoint, reusing or transferring information from previously learned tasks for the learning of new tasks has the potential to significantly improve the sample efficiency of a reinforcement learning agent.

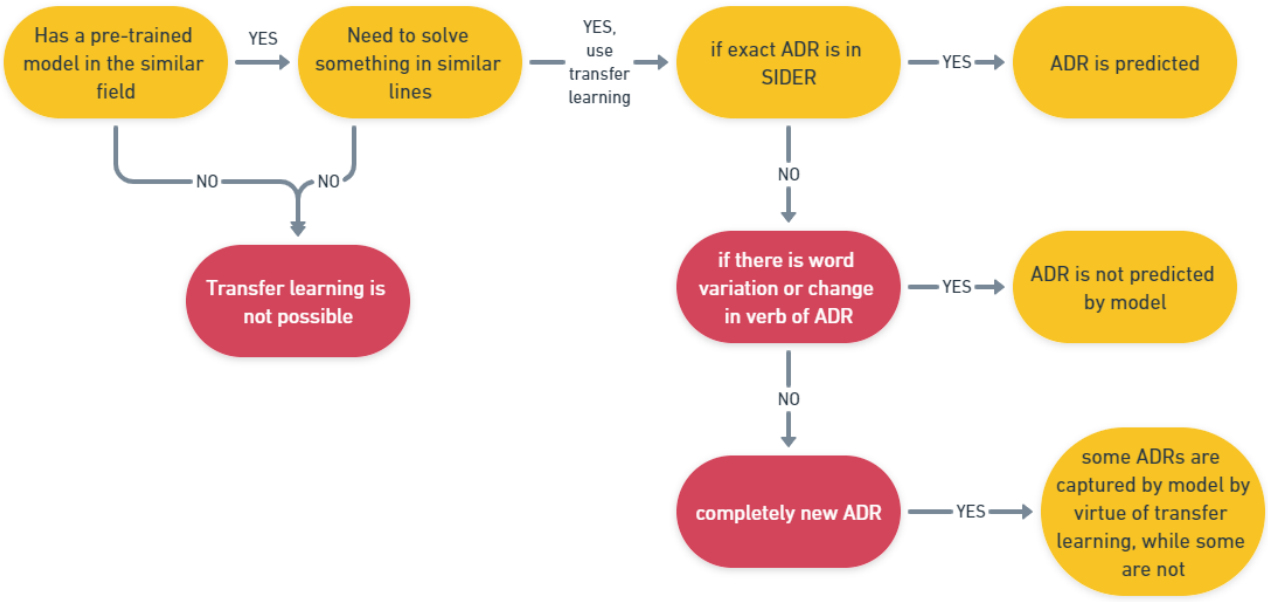

Figure 1: A decision tree explaining how transfer learning works for our model.

This learning method is an application of deep learning, which implies it demands more data for training purposes and consequently, more time of computation. In this project, we have analysed transfer learning on the pre-trained BioBERT model and tested over inter-drug data and inter-website data. In figure 1, we explicitly show how our transfer learning framework works, processes and how differently it can predict outcome than usual.

3 Proposed Framework

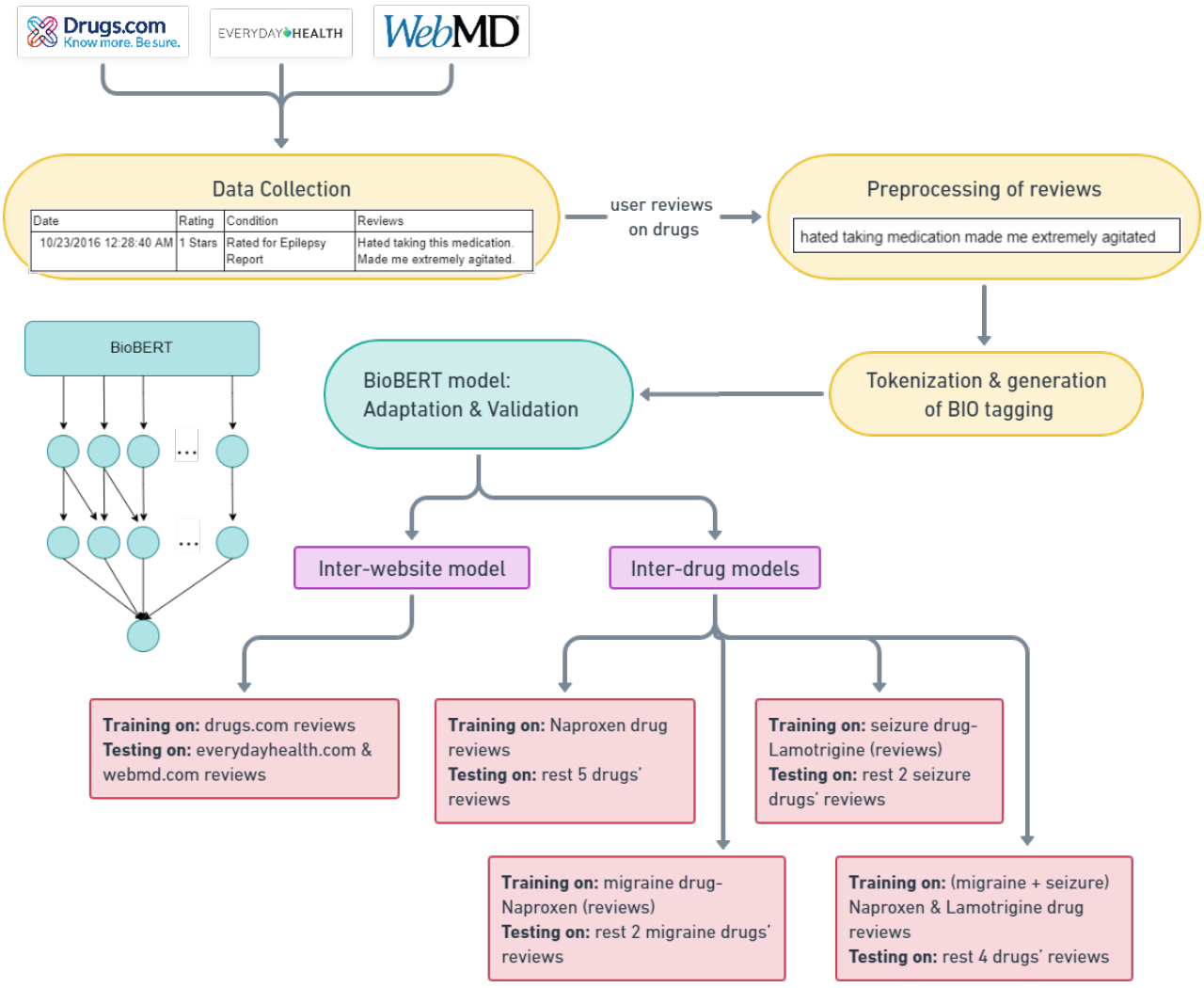

Figure 2: Architectural framework of proposed model.

The proposed model is explicitly shown in figure 2. First we extract data from the websites mentioned in Section 3.1. The data is then processed so that it is apt for further processes, explained in Section 3.3. Next in Section 3.4, we explain how the reviews are tokenized word by word and BIO tags are generated for each word. Section 3.5 explains how our model adapts to the existing solution and validates on our drugs' review data, thus leading to five different models.

3.1 Data Collection

For our model, we extracted user reviews with various other attributes like rating, date, time, age, gender etc from the following three websites. Each of these websites are well-

known in medical domain for authenticity of their data and is popular among researchers as their data is available for scrapping and data mining purposes.

- www.webmd.com
- www.everyayhealth.com
- www.drugs.com

In addition to the above websites, we had another website- www.askapatient.com, with good reviews that were apt for text mining. However, they did not allow web or data scrapping from their websites. While another website- www.druglib.com, though had elaborated user reviews, had very less number of reviews for each drug in the specific categories of condition.

Each drug may be used in several medicine in some proportion. Thus, when we search reviews for a specific drug, the context of treatment differs. In this case, we have sorted our data according to two medical disorders, i.e. migraines and seizures, their types and some their closely-related causes.

The dataset obtained consists of the following attributes, shown in Table 1, from the three different websites:

Table 1: Data sample from each website.

Website	Attributes	Sample
drugs.com	Date	May 13, 2019
	Medication duration	Taken for 6 months to 1 year
	Condition	For Schizoaffective Disorder:
	Review	Carbamazepine has been a miracle for schizoaffective disorder
	Rating	9
everydayhealth.com	Date	10/23/2016 12:28:40 AM
	Rating	1 Stars
	Condition	Rated for Epilepsy Report
	Review	Hated taking this medication. Made me extremely agitated.
webmd.com	Date	10/30/2007
	Age	45-54
	Condition	Head Pain
	Review	I have migraine headaches and acetaminophen oral just doesn't cut it for my type of headaches.
	Ease of use	5
	DrugID	362
	Effectiveness	1
	Satisfaction	1
	Sex	Female
	Listed side effects	This drug usually has no side effects.
	Usefulness	2

3.2 Data Statistics

In the table below, we have listed the number of user reviews that we collected from the three websites for the mentioned drugs.

Table 2: Data statistics.

	everydayhealth.com	**webmd.com**	**drugs.com**
Acetaminophen	141	136	89
Carbamazepine	188	150	313
Lamotrigine	156	72	155
Levetiracetam	241	117	380
Naproxen	211	742	637
Sumatriptan	372	152	507
Total	1309	1369	2081

3.3 Pre-processing of Data

The drug reviews that we have obtained contain noise of various forms. Text preprocessing is required to clean the data and make it ready to feed the model. In this process, we perform the following on our text data (reviews):

1. removal of email IDs
2. removal of emojis
3. removal of url, links
4. removal of punctuation marks and special characters
5. lower-casing of text
6. removal of stopwords

In the following Table 3, we show a sample of the processed text from the original text reviews which is further analysed in our model.

Table 3: Pre-processed text reviews.

	Original review	**Processed review**
1.	Gives terrible stomach pain after only a few uses.	*gives terrible stomach pain after only a few uses*
2.	I have no seizures since starting this mediation.	*i have no seizures since starting this mediation*
3.	I never have been so sick in my life after taking one pills. Stomach pains, heart burn, nausea, body aches. I thought I was going to die. Be careful of this drug! Naproxen 500mg tab.	*i never have been so sick in my life after taking one pills stomach pains heart burn nausea body aches i thought i was going to die be careful of this drug naproxen 500mg tab*
4.	I had been taking naproxen for 15 years my kidneys are damaged l don't know way my Doctor kept me on this drug if it was going to cause this kind of damage to my body is there a lawsuit pending in the courts for Naproxen.	*i had been taking naproxen for 15 years my kidneys are damaged l dont know way my doctor kept me on this drug if it was going to cause this kind of damage to my body is there a lawsuit pending in the courts for naproxen*

3.4 Tokenization & BIO Tagging

Tokenization is the process of breaking the raw text data into smaller chunks or into words, called tokens. This is an important step in NLP due to many reasons. It helps in understanding the context of data or for developing the model for NLP. The tokenization process helps in interpreting the meaning of the text by analyzing the sequence of the words. The token occurrences in each review are then mapped to their corresponding token IDs as a vector representing that review. Consider the following example, the preprocessed text is tokenized into words initially.

this *medicine* *caused* *breathing* *difficulty* *and* *swelling* *in* *my* *body*

[this, medicine, caused, breathing, difficulty, and, swelling, in, my, body]

This is followed by numericalization where token IDs are given to the corresponding tokens. We map each token to a unique integer in the corpus' vocabulary as follows:

[this,	medicine,	caused,	breathing,	difficulty,	and,	swelling,	in,	my,	body]
[1143,	5183,	2417,	4944,	7263,	1106,	20086,	1108,	1140,	1405]

Further, BIO tags are generated for each tag by referring to the SIDER dataset we obtained. This is shown as follows:

this	*medicine*	*caused*	*breathing*	*difficulty*	*and*	*swelling*	*in*	*my*	*body*
O	O	O	B	I	O	B	O	O	O

Here, the words are tokenized and tagged as labels B, I and O where B is tagged to the first word of an ADR and all sequel parts of the same ADR are tagged as I, while all of the non-ADR terms are tagged O.

The BIO or IOB format (short for inside, outside, beginning) is a tagging format for tagging tokens in a chunking task in computational linguistics. The B- prefix before a tag indicates that the tag is the beginning of a chunk, and an I- prefix before a tag indicates that the tag is inside a chunk. The B- tag is used only when a tag is followed by a tag of the same type without O tokens between them. An O tag indicates that a token belongs to no entity/chunk. We use this format to tag our training data for accomplishing the objective of named entity recognition of ADRs.

We use the SIDER (SIDe Effect Resource) dataset for tagging of our training dataset. SIDER contains a listing of side effects extracted from Medical Dictionary for Regulatory Activities (MedDRA).

3.4.1 Problems experienced during labelling the dataset:

The following sample review shows the problem we faced while labelling a sample string: *"I have back pain and pain from taking this medicine"*

In the above sentence, the ADRs according to SIDER are as follows:
["back pain", "pain"]
The following problem arises when labelling the data: Initially, we label "back pain" in our string:

I	*have*	*back*	*pain*	*and*	*pain*	*from*	*taking*	*this*	*medicine*
O	O	B	I	O	O	O	O	O	O

However, when we label "pain" in our string, we got the following sequence of labels:

I	*have*	*back*	*pain*	*and*	*pain*	*from*	*taking*	*this*	*medicine*
O	O	B	B	O	B	O	O	O	O

This arises due to the fact that "pain" in itself is an ADR, as cited by SIDER.

Solution:

In order to tackle this problem, we labelled the data in accordance with the number of words an ADR has, starting from the ADRs with one word and going all the way up to ADRs having n words, where n is the maximum number of words an ADR can contain. Taking the above example as a test case, after sorting the list based on the number of words each ADR has, we get the following updated ADR list: ["pain", "back pain"] When we label the string with "pain", we get:

I	*have*	*back*	*pain*	*and*	*pain*	*from*	*taking*	*this*	*medicine*
O	O	O	B	O	B	O	O	O	O

After labelling the string with "back pain", we get:

I	*have*	*back*	*pain*	*and*	*pain*	*from*	*taking*	*this*	*medicine*
O	O	B	I	O	B	O	O	O	O

3.5 Model Adaptation & Fine-tuning

By using transfer learning method, we adapt the BioBERT model to our dataset. BioBERT is the pre-trained model, where we optimize our parameters and weights of the model, taking reference of the BioBERT model. We fine-tune our model in context of drug reviews that is needed for our purpose. We, thus, form the following models for the purpose of analysing different characteristics of transfer learning for detection of ADRs from users' reviews on drugs.

- **Inter-website model:** Training on all drugs' reviews of drugs.com website, testing on all drugs' reviews of everydayhealth.com and webmd.com websites.
- **Inter-drug models:**
  - Training on Naproxen drug reviews, testing on rest five drugs' reviews.
  - Training on migraine drug- Naproxen drug reviews, testing on rest two migraine drugs' reviews.
  - Training on seizure drug- Lamotrigine drug reviews, testing on rest two seizure drugs' reviews.

– Training on both migraine and seizure drugs- Naproxen & Lamotrigine drugs' reviews, testing on rest four drugs' reviews.

4 Experimental Set-up & Results

For the fine-tuning of models mentioned in Section 3.5, we optimize the parameters to enable the models to perform the task of detecting ADRs with relative accuracy. Consequently, Section 4.1 gives a tabular view of all parameters for each model and the subsequent sections give us the resluts of the same.

4.1 Parameter Setting

Each model has 4 parameters which are the properties of a training model, that can be varied for it to perform best during the learning process. Our model consists of four parameters, namely, learning rate which is varied from $(1e-5, 1e-1)$, weight decay that is varied from $(1e-5, 1e-1)$, drop-out that is varied between $(0.2, 0.4)$ and epochs which is varied from $(10, 100)$. The final parameters selected are shown in Table 4 against each model.

Table 4: Parameters set for all models.

	Learning Rate	**Weight Decay**	**Drop-out**	**Epochs**
Inter-website model:	1.4458369832358114e-05	8.719494418047482e-05	0.3	50
Inter-drug model:				
Trained on Naproxen	0.0001	0.0001	0.3	15
Trained on migraine drug: Naproxen	0.0001	0.0001	0.3	15
Trained on seizure drug: Lamotrigine	0.0001	0.0001	0.3	15
Trained on both: Naproxen & Lamotrigine	0.0001	0.0001	0.3	15

4.2 Inter-Website Model

Model 1:

This model is based on transfer learning from the BioBERT model where we combine

drug reviews of all drugs, i.e. of both migraine and seizure drugs, of a single website- www.drugs.com. We test this model on the combined drugs reviews of all drugs of www.everydayhealth.com and that of www.webmd.com. The following Table 5 gives us the validation majors of training data, while Table 6 and Table 7 gives us the test results on the respective other two websites, for B, I and O separately.

Table 5: Validation majors on training data of all drug reviews of www.drugs.com website.

		Precision	Recall	F1-score	Support
drugs.com	B	1.00	0.99	1.00	9696
	I	0.98	0.99	0.98	1308
	O	1.00	1.00	1.00	267393

Table 6: Test results on data of all drug reviews of www.everydayhealth.com website.

		Precision	Recall	F1-score	Support
everydayhealth.com	B	0.97	0.92	0.94	5978
	I	0.89	0.79	0.84	724
	O	1.00	1.00	1.00	188780

Table 7: Test results on data of all drug reviews of www.webmd.com website.

		Precision	Recall	F1-score	Support
webmd.com	B	0.96	0.94	0.95	2252
	I	0.91	0.80	0.85	271
	O	1.00	1.00	1.00	53143

4.3 Inter-Drug Models

Model 2:

In this model, we choose the dataset of the drug having maximum reviews- Naproxen drug, out of all the six drugs for training. We then test this model on the drugs reviews of all the other drugs, namely, Acetaminophen, Carbamazepine, Lamotrigine, Levetiracetam and Sumatriptan. Each dataset of a specific drug consists of drug reviews of that drug from all the three websites. The following Table 8 gives us the validation majors of training data, while Table 9 gives us the test results on the other drugs, for B, I and O separately.

Table 8: Validation majors on training data of Naproxen drug reviews of all websites.

		Precision	Recall	F1-score	Support
Naproxen	B	0.95	0.92	0.92	937
	I	0.83	0.81	0.82	291
	O	0.99	1.00	1.00	17294

Table 9: Test results on drug reviews data of Acetaminophen, Carbamazepine, Lamotrigine, Levetiracetam and Sumatriptan of all websites.

		Precision	Recall	F1-score	Support
Acetaminophen	B	0.91	0.93	0.92	824
	I	0.75	0.74	0.74	61
	O	1.00	0.99	1.00	15936
Carbamazepine	B	0.78	0.80	0.79	1332
	I	0.79	0.62	0.69	239
	O	0.99	0.99	0.99	28412
Lamotrigine	B	0.77	0.66	0.71	806
	I	0.50	0.17	0.26	167
	O	0.99	0.99	0.99	14573
Levetiracetam	B	0.78	0.59	0.67	1381
	I	0.55	0.08	0.14	283
	O	1.00	1.00	0.99	56417
Sumatriptan	B	0.94	0.94	0.94	3375
	I	0.68	0.46	0.55	133
	O	1.00	1.00	1.00	75994

Model 3:

In this model, we choose the dataset of the migraine drug having maximum reviews-Naproxen drug, out of all the three migraine drugs for training. We then test this model on the drugs reviews of all the other migraine drugs, namely, Acetaminophen and Sumatriptan. Each dataset of a specific drug consists of drug reviews of that drug from all the three websites. Since the training data here is same, considering the validation majors of training data in Table 8, we have the test results on the other migraine drugs in Table 10, for B, I and O separately.

Table 10: Test results on migraine drug reviews data of Acetaminophen and Sumatriptan of all websites.

		Precision	Recall	F1-score	Support
Acetaminophen	B	0.91	0.93	0.92	824
	I	0.75	0.74	0.74	61
	O	1.00	0.99	1.00	15936
Sumatriptan	B	0.94	0.94	0.94	3375
	I	0.68	0.46	0.55	133
	O	1.00	1.00	1.00	75994

Model 4:

In this model, we choose the dataset of the seizure drug having maximum reviews- Lamotrigine drug, out of all the three seizure drugs for training. We then test this model on the drugs reviews of all the other seizure drugs, namely, Carbamazepine and Levetiracetam. Each dataset of a specific drug consists of drug reviews of that drug from all the three websites. The following Table 11 gives us the validation majors of training data, while Table 12 gives us the test results on the other two seizure drugs, for B, I and O separately.

Table 11: Validation majors on training data of Lamotrigine drug reviews data of all websites.

		Precision	Recall	F1-score	Support
Lamotrigine	B	0.96	0.92	0.94	1381
	I	0.87	0.83	0.85	283
	O	0.99	1.00	1.00	56417

Table 12: Test results on seizures drug reviews data of Carbamazepine and Levetiracetam of all websites.

		Precision	Recall	F1-score	Support
Carbamazepine	B	0.85	0.85	0.85	1332
	I	0.81	0.64	0.71	239
	O	0.99	1.00	1.00	45257
Levetiracetam	B	0.83	0.85	0.84	806
	I	0.82	0.59	0.68	167
	O	0.99	1.00	0.99	29574

Model 5:

In this model, we choose the dataset of the both migraine & seizure drugs having maximum reviews, i.e. Naproxen and Lamotrigine drugs, for training. This model is then tested on the drugs reviews of all the other drugs, namely, Acetaminophen, Carbamazepine, Levetiracetam and Sumatriptan. The following Table 13 gives us the validation majors of training data, while Table 14 gives us the test results on the rest four drugs.

Table 13: Validation majors on training data of Naproxen & Lamotrigine drug reviews data of all websites.

		Precision	Recall	F1-score	Support
Naproxen & Lamotrigine	B	0.94	0.90	0.92	2318
	I	0.84	0.82	0.83	574
	O	1.00	1.00	1.00	73711

Table 14: Test results on other four migraine & seizures drug reviews data- Acetaminophen, Carbamazepine, Levetiracetam and Sumatriptan of all websites.

		Precision	Recall	F1-score	Support
Acetaminophen	B	0.95	0.91	0.93	824
	I	0.81	0.79	0.80	61
	O	0.99	1.00	1.00	15936
Carbamazepine	B	0.87	0.71	0.78	1332
	I	0.89	0.63	0.74	239
	O	0.99	1.00	0.99	45257
Levetiracetam	B	0.79	0.46	0.58	1381
	I	0.93	0.05	0.09	283
	O	0.98	1.00	0.99	56417
Sumatriptan	B	0.95	0.93	0.94	3375
	I	0.90	0.46	0.61	133
	O	1.00	1.00	1.00	75994

5 Result Analysis

With the above models and experimental results, we have a couple of observations and analysis. The Table 15 gives us an example of how the ADRs that exist in SIDER are captured by the model, with respective BIO-tags of each token.

Table 15: ADRs labelled with SIDER and captured by the model.

Text																**SIDER ADRs**	**Model ADRs**
tramadol	the	only	side	effect	is	constipation										constipation	constipation
O	O	O	O	O	O	B											
forgot	to	mention	i	experience	dry	mouth	acid	reflux	and	ibs	bummer	but	the	drug	is	effective	dry mouth	dry mouth
O	O	O	O	O	B	I	O	O	O	O	O	O	O	O	O	O		
it	is	working	very	well	i	did	have	an	aweful	headache	for	the	first	5	days		headache	headache
O	O	O	O	O	O	O	O	O	O	B	O	O	O	O	O			

On the other hand, Table 16 shows examples of those ADRs that are not in the SIDER but captured by the model. Indications not in SIDER as ADRs are caught by the model because BioBERT is trained on millions of biomedical text data from PubMed and PMC and therefore, has learned to detect the various named entity present in those text.

Table 16: ADRs not in SIDER but captured by the model.

Text	**SIDER ADRs**	**Model ADRs**
the only side effect he is experienced so far is sleepiness O O O O O O O O O O B		sleepiness
he has a little crankiness but that could be from being two O O O O B O O O O O O O		crankiness
i take 1000 mgs at bedtime and have developed muscle cramps in the ankles and calves O O O O O O O O O B I O O O O O		muscle cramps

Table 17 cites instances showing how the model captures not just ADRs but severity as well to a great extent.

Table 17: Indications caught by the model.

Text	**SIDER ADRs**	**Model ADRs**
when a chronic headache comes on i inject myself 3 to 4 minutes later no pain O O B I O O O O O O O O O O O O O	headache, pain	chronic headache
they managed to break my wrist also i suffered chronic stress from this O O O O B O O O O B I O O	stress	chronic stress
i was suffering from chronic fatigue anyway O O O O B I O	fatigue	chronic fatigue

Some ADRs in the SIDER are not caught by the model because the model is trained with limited amount of dataset comprising only few hundred of reviews. This limits the model from learning the contextual abstracts of the dataset pertaining to the problem of ADR extraction and therefore, the model is unable to extract ADRs from the dataset. This is shown in Table 18.

Table 18: ADRs not caught by the model.

Text														**SIDER ADRs**	**Model ADRs**
i	no	longer	take	this	medication	and	no	longer	side	effects	except	for	dystonia	dystonia	
O	O	O	O	O	O	O	O	O	O	O	O	O	B		
trileptal	seems	to	be	starting	to	give	me	migraines	too					migraine	
O	O	O	O	O	O	O	O	B	O						
it	gave	me	mood	swings										mood swing	
O	O	O	B	I											

In Table 19, we analyse the top 10 ADRs of each drug that occur in our results of inter-drug models. Their corresponding frequency of occurence is also mentioned immediately below the ADR.

Table 19: ADRs frequency (%) for each drug on analysing results of Table 14.

Drug	**Top 10 ADRs frequency (%)**									
Acetaminophen	nausea	dizziness	constipation	dry mouth	drowsiness	swelling	anxiety	chills	insomnia	stress
	17.91	8.96	8.96	7.46	5.97	4.48	2.99	2.99	2.99	2.99
Carbamazepine	dizziness	rash	nausea	drowsiness	headache	fatigue	unwell	sleepiness	migraine	dry mouth
	6.48	6.48	5.67	4.45	4.05	3.64	2.43	2.02	2.02	1.62
Levetiracetam	depression	anxiety	fatigue	pain	irritability	dizziness	insomnia	crying	rash	sleepiness
	20.11	15.58	5.38	5.10	4.53	4.25	3.68	2.27	1.98	1.70
Sumatriptan	nausea	vomiting	dizziness	burning sensation	drowsiness	chest pain	fatigue	numbness	anxiety	sleepiness
	23.86	12.06	4.02	3.75	3.49	2.95	2.68	1.88	1.88	1.88

In Table 20, we analyse how many of the top 10 ADRs match with the published common or severe ADRs or those ADRs developing due to overdose of the drug. We refer government sites and trusted medical domain sites like `https://medlineplus.gov/` and `https://pubmed.ncbi.nlm.nih.gov/`. Here, we note that the rest of the ADRs that are detected are genuine, however, it may be such that only a small percentage of the patients suffer from these ADRs as a result of various cases like that of overdose, or allergen effect or a typical exception.

Table 20: List of new ADRs out of top 10 ADRs, predicted by model.

Drug	**Count of:**		**Rare ADRs**
	matched known ADR	**rarely occuring ADR**	
Acetaminophen	8	2	anxiety, insomnia
Carbamazepine	9	1	drowsiness, sleepiness
Levetiracetam	8	2	depression, insomnia
Sumatriptan	9	1	anxiety

6 Conclusion & Future Scope

Transfer learning to detect ADRs from user reviews using BioBERT leverages us to understand how models can be enhanced for different learning tasks from a solved model. Our model was successful in understanding the context and not only bring forth legitimate ADRs that the SIDER does not contain, but also capture the severity to an extent. The model also showed that word variations and change of verb of the ADRs needs to be taken care of additionally. Our experimentation with the model enables us the conclude that such deep learning models are data demanding and may not work as desired for small amount of data. Thus, we chose the largest datasets in each category of model that we performed for training purposes. We have an analysis of the drug reviews of the chosen six drugs in our case, to what extent transfer learning can handle different tasks in different contexts of data, and how it can detect ADRs. Another analysis leads us to the conclusion that aggregating heterogeneous drugs' reviews increases the effectiveness of our work and thus, the cross-domain detection of ADRs comes with more acceptable results. Including more domain knowledge enhances the model furthermore.

With the current analysis of data in hand, we can further extend the work to find correlation between ADRs with respect to age or gender. The varying ADR resulted according to the group of patients in different age limits may be an interesting anecdote. Moreover, the gender biased results, like hormonal effect in females or males, may also lead to interesting results.

System Specifications

The above tools are developed and executed in the following systems of these specifications mentioned below:

Table 21: Specifications of systems used.

	System 1	**System 2**	**System 3**
Manufacturer	Lenovo	Acer	HP
Processor	Intel(R) Core(TM) i7-4700MQ CPU @ 2.40GHz	Intel(R) Core(TM) i3-5005U CPU @ 2.00GHz	Ryzen5 4600h 3.00GHz
RAM	8.00 Gb	4.00 Gb	16.00 Gb
System Type	64-bit Operating System, x64-based processor		
IDE	Google Colab		

References

[1] Cristóbal Colón-Ruiz and Isabel Segura-Bedmar. Comparing deep learning architectures for sentiment analysis on drug reviews. *Journal of Biomedical Informatics*, 110:103539, 2020.

[2] Vinodhini Gopalakrishnan and Chandrasekaran Ramaswamy. Patient opinion mining to analyze drugs satisfaction using supervised learning. *Journal of applied research and technology*, 15(4):311–319, 2017.

[3] Felix Gräßer, Surya Kallumadi, Hagen Malberg, and Sebastian Zaunseder. Aspect-based sentiment analysis of drug reviews applying cross-domain and cross-data learning. In *Proceedings of the 2018 International Conference on Digital Health*, pages 121–125, 2018.

[4] Sajid Hussain, Hammad Afzal, Ramsha Saeed, Naima Iltaf, and Mir Yasir Umair. Pharmacovigilance with transformers: A framework to detect adverse drug reactions using bert fine-tuned with farm. *Computational and Mathematical Methods in Medicine*, 2021, 2021.

[5] Trung Huynh, Yulan He, Alistair Willis, and Stefan Rüger. Adverse drug reaction classification with deep neural networks. Coling, 2016.

[6] Chun Yen Lee and Yi-Ping Phoebe Chen. Prediction of drug adverse events using deep learning in pharmaceutical discovery. *Briefings in bioinformatics*, 22(2):1884–1901, 2021.

[7] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, 2020.