Journal of Theoretical and Applied Information Technology
29th February 2024. Vol.102. No 4

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

LEARNING DISTRIBUTED REPRESENTATION OF DRUG SEQUENCES FROM ADVERSE EVENT REPORTING DATA

RASHA ASSAF[1], AMJAD RATROUT[2], MOHAMMED KHALILIA[3] , RASHID JAYOUSI[1]

[1]Al-Quds University, Department of Computer, Jerusalem, Palestine

[2]Arabic-American University, Department of Computer, Jenin, Palestine

[3]Birzeit University, Department of Computer, Birzeit, Palestine

E-mail: rasha.assaf@students.alquds.edu, amjad.rattrout@aaup.edu, mkhalilia@birzeit.edu, rjayousi@staff.alquds.edu

ABSTRACT

Predicting adverse event reactions (ADR) is challenging as it depends on the patient's condition, pre-existing conditions, and the different medications being administered. One source of data for ADR that we believe is under-utilized is the FDA Adverse Event Reporting System (FAERS) electronic submissions. FARES contains a large number of ADRs including drugs and their attributes (timestamp, dosage, route, duration, etc.), in addition to reactions and outcomes. In this paper, we utilize FARES data to model each ADR as a sequence of medications to train a model that learns the similarity between the drug sequence and the ADRs. As a by-product of this work, we also learn the drug sequence representations, which can be used for other down streaming tasks. Our model is based on a transformer to encode drug sequences and adverse events, the model then outputs the likelihood of the ADR given the drug sequence. Our best model achieved 0.87 F1 score, showing efficient representations for the drug sequences. We also performed qualitative analysis to validate the drug sequence representations. To our knowledge, we are the first to utilize drug sequences from FARES data to learn drug embeddings.

Keywords: *Adverse Events, Embeddings, Sequence Modeling, LSTM, Transformers, RoBERTa*

1. INTRODUCTION

As defined by the World Health Organization (WHO)," an adverse drug reaction (ADR) is typically referred to as an unanticipated and harmful response that is thought to have been brought on by a medicine taken as directed" [1]. It is widely known that the prevalence of ADRs poses a serious global public health risk. Over two million serious ADRs occur in hospitals each year in the United States, amounting to over 100,000 fatalities [2]. Early detection of probable ADRs linked to drug candidates during the initial phases of drug development can enhance drug safety, diminish risks for patients, and generate cost savings for pharmaceutical corporations [3]. In summary, understanding the overall risk-benefit profile of new medicines, maintaining patient safety, and making informed regulatory decisions all depend on the research of adverse events in clinical trials. It supports the proper introduction of novel treatments into clinical practice and the moral conduct of clinical research.

According to several reports, there have been instances of drug recalls after their approval by the Food and Drug Administrator (FDA) [4]. For example, in January 2022, concerns regarding the presence of N-nitrosodimethylamine (NDMA), a chemical that may be carcinogenic, led to recall and withdrawal of Ranitidine, known by the trade name Zantac. Based on lab testing, NDMA is categorized as a potential human carcinogen. Between 1953 and 2014, 43 medications were discontinued globally, and 462 medicines were withdrawn from the United States market [5]. According to the FDA website, a total of 113 drugs that had received approval between 2015 and 2017 were taken off the market after 2017. For instance, Vioxx, which treats arthritis chronic pain gained FDA approval in 1999 but was withdrawn from the market five years later due to an increased risk of stroke and heart attack [6].

To alleviate this issue, researchers started looking into ways to predict ADRs ahead of time. Some of the techniques proposed include rule-based, statistical, machine learning, and hybrid approaches. Some of the techniques can consume

a ⅄considerable amount of time, are not cost-effective, and rely heavily on domain experts and feature engineering. Deep learning approaches [7, 8, 9], on the other hand, have shown significant improvement in many areas including sequence modeling. In this paper, we utilize deep learning approaches to model drug sequence information in the context of adverse event reaction prediction.

In this paper, we propose a framework for learning drug sequence representations and their corresponding adverse events. Our approach employs a transformer encoder to encode both the drug sequence and ADR in the FARES data. The goal of the model is to learn the likelihood of an ADR given a drug sequence and to learn drug sequence representations.

The rest of this paper is organized as follows: Section 2 covers related work, Section 3 introduces the FARES dataset, Section 4 describes the methodology, Section 5 presents the results, and we conclude in Section 6.

2. RELATED WORK

Numerous works published on ADR prediction, but very few utilized the FARES data. The most relevant published work is the aer2vec [10], using Proportional Reporting Ratio (PRR) the extent to which a particular adverse event is reported for individuals taking a specific drug, compared to the frequency at which the same adverse event is reported for patients taking some other drug. They used the drug name and to improve the performance use included the drug susceptibility (primary or secondary suspect). The Area Under the Curve based on the OMOP is 0.646 compared to FARES of Area Under the Curve (AUC) 0.744.

The authors in [11] aim to provide a comprehensive review of the application of deep learning techniques in detecting ADRs caused by single drugs and drug-drug interactions (DDIs). The paper offers an in-depth analysis of various deep learning approaches employed in each case, and strives to enhance the understanding of deep learning concepts and their advancements in extracting ADRs. The authors emphasize the significance of incorporating diverse data sources and employing novel deep learning techniques to detect ADRs. Deep learning techniques have been applied to detect ADRs brought on by both single medications and DDI.

The authors in [12] obtained information about associations between drugs and ADRs from the SIDER [13] database. They created ten different types of chemical fingerprints using molecular structures. For each fingerprint, they created logistic regression models with L2 norm regularization to predict ADRs. Additionally, they combined the creation of neural fingerprints and the creation of models using a convolutional deep learning framework. All eleven models' performance were assessed, and the results showed that the neural fingerprints had the best overall performance.

The authors in [14] focused on using network-based methods to predict drug-target interactions (DTI). Specifically, they developed a recommendation algorithm-based network-based inference (NBI) approach. They first discussed the approaches and assessments of network-based techniques before highlighting some of their many applications in various fields. Target prediction and the investigation of the molecular mechanisms underlying therapeutic effects or safety issues were included in these applications

The authors in [15] developed a graph neural network design that achieves cutting-edge results in predicting probable side effects from combination drug use. Their work is entirely dependent on the molecular structure data of these drug pairs. They demonstrated the efficiency of combining comprehensive drug-drug insights while creating representations for individual drugs through the process of message propagation within each drug and concurrent attention to the structure of the other drug.

In [16], the authors presented a prediction algorithm to identify instances of probable adverse event underreporting. The authors simulated the effect on model performance in cases of underreporting of adverse events. The model has an AUC of 0.62, 0.79, and 0.92 reflecting 25%, 50%, and 75% underreporting of adverse events, respectively.

In [17], the authors have formulated an adaptable graph convolutional method capable of incorporating molecular influences stemming from the varying count of medications commonly consumed by an average patient. Their model surpasses conventional machine learning techniques in prognosticating hospitalization and mortality within the UK Biobank dataset, showcasing an R square value of 0.37 and an AUC score of 0.90.

3. FAERS DATASET

The FAERS is a database used for the voluntary reporting of adverse events of drugs and therapeutic biological products (including vaccines) [4]. FAERS includes data covering United States market products and reports from all over the world. It also includes patient demographics and segments (elderly, children, and pregnant women), which makes it the largest database in pharmacovigilance for researchers. However, FAERS data lacks standards and naming convention. That is because reports come from different regions in the world and the data is entered in free text format and does not use any standard terminologies such as RxNORM or SNOMED. This causes problems such as typographical mistakes when entering medication names. Moreover, the FAERS database, as well as any other automatic adverse reporting system, have repeated reports as a concern. So, with the text- processing procedures that will be used with database entries, especially drug names entries, it's necessary to conduct length and resource-intensive database analysis. This will improve the quality of the data and it will allow the use of common vocabulary [18]. The data used in this work covers the year 2021 and includes 1,860,388 cases and 6,331,148 drugs (count of unnormalized drug names and typos inflate the count significantly). In addition, FARES data contains demographic data (persons age, gender, and location at the time of the adverse event), medical information (time of ADR, the nature of the incident, side effect or medication error, outcomes such as hospitalization or death), product details (name, dosage form, administration method, active components of the specific medication, biologic, or medical device that was linked to the adverse event), information about the person or organization that reported the adverse event to the FDA, which can include consumers, manufacturers, and healthcare providers and patient outcome (death, hospitalization, injury, etc.). There are over 15K reactions in our dataset, and similar to the drugs, this represents the count of unnormalized ADRs and typos can significantly inflate that number. There is an average of 20 reactions per sequence, with a total set of 800K drug-reaction sequence pairs. Hence, modeling the problem as multiclass classification will be challenging, even if we used a sparse network between the last hidden layer and the output layer. Likely, since the reactions do not have an explicit ordinal relationship nor a mutual effect that can be drawn from their context (drug sequence), the problem cannot be modeled as a sequence-to-sequence task.

Example case from the FARES data is shown in Figure 1, which shows the sequence drugs taken over time. Other information we know about this case includes patient demographics (female, 61 years old and located in Canada), adverse event reactions (Arthralgia, Blood cholesterol increased, Bronchitis, Dizziness, Hypotension, Rheumatoid arthritis) and the outcome, which is hospitalization.

3.1 Data Preprocessing

The drug name information encoder uses a byte-pair-encoding (BPE) tokenizer. Hence, clean-up and normalization of the drug names were necessary to ensure efficient tokenization. First, special characters are removed from the names, except for the hyphens (-) and dots (.) which are replaced by a single whitespace. Then, we x the mis-concatenation of drugs which have multipart names. For example, Esmolol Hydrochloride is fixed to be esmolol hydrochloride. We propose an iterative algorithm1 to fix this miswording. The algorithm starts by building a set of drug names separated by spaces, then iteratively builds subset S whose elements do not have sub-tokens from other elements.

Algorithm 1: Drugs Misspelling Preprocessing

Input: D = Drugs Sequences Concatenated
Output: S = { s | $s \in S \wedge s \notin R$ }

1: *Set* $R \leftarrow split(D, "/")$
2: *while* $R \neq \phi$ *do*:
3: $S \leftarrow \{r \mid r \in R \wedge \{\neg \exists x \in R \ substoken(r, x) = True\}\}$
4: $R \leftarrow \{r \mid r \in R \wedge r \notin S\}$
5: *if* $R \neq \phi$:
6: $R \leftarrow splitSupertokens(R, S)$
7: *end while*

Algorithm1

The remaining drug names are added to set R. Finally, drug names in set R are split using the elements of S. Iterations stop when R is an empty set. The final set S is finally used to split the text into drug names. The final preprocessing step is to lowercase the text.

4. METHODOLOGY

We propose an end-to-end training framework for learning the drug sequence representations. We frame our approach as a sequence labeling problem; given a sequence of drugs, we predict whether the ADR is associated with the drug sequence or not. In this section, we will explain each component of our model.

4.1 Notations Concept

Before presenting our methodology, we will set the notation. The FARES data consists of a sequence of drugs that are time stamped, we refer to this sequence as S_i, such that $S_i = (s_1, s_2, \ldots, s_n)$ and s_j, refers to the jth drug in the sequence. For each S_i there are corresponding list of ADRs, which we denote as R_i such that $R_i = \{r_1, r, \ldots, r_m\}$ and r_z, refers to the zth ADR. Both S_i and D_i form a case in the FARES dataset, $c_i = \{S_i, D_i\}$. In order to associate each ADR, r_z, with the drug sequence S_i, we unfold the ADRs and form new tuples of sequence-reaction, $c_{ij}^p = \{S_i, r_{ij}\}$ for each $0 \leq j \leq m$. The case c_{ij}^p represents a positive example. We also generate negative examples, where k-negative samples are selected using hard negative mining and added to the data records, we refer to the negative example as c_{ij}^n. Hence, the model is treated as a sequence classification one. The objective is to maximize the likelihood $P(y = 1 \mid S_i; r_z; \theta)$, the probability of seeing the reaction r_z for a given drug sequence S_i, parameterized by model weights θ.

4.2 Hard Negative Sampling

The negative sampling technique is critical for the efficiency of learning representation vectors by discrimination [19]. Negative samples in our problem are reactions that do not appear with the associated drugs sequence, hence labeled as 0. Hard negative samples are negative points, here reactions, that are difficult for the model to differentiate from a true one. For each drugs sequence S, reactions sampled as hard negatives should have at least 0.65 average cosine similarity, using their respective representation vectors, to other reactions of the sequence S. Reactions satisfying this condition are then sorted ascendingly by the frequency by which they occur with each of the drugs in the sequence S. Finally, a set of at most three hard negative reactions for each true reaction are selected to be among the total negative samples.

4.3 Model Architecture

The model overall architecture is summarized in Figure 2. The model takes two types of inputs derived from the FAERS data. The drug sequence as depicted in Figure 1 and for this type of input we only use the drug names. The second input is the ADR, note that there can be multiple ADRs per drug sequence, and as mentioned in Section 4.1, we unfold the ADRs for each drug sequence to form the input, which is one ADR per drug sequence. Each type of input is fed through its encoder to generate representations. We experimented with different types of encoders for the drug sequence encoding as we will explain later. At this stage, we have one representation of the drug sequence and one for the ADR, which are then concatenated and used as input to the projection layers. The final step computes the softmax to determine if the drug sequence causes the ADR or not.

4.4 Drug Sequence Encoder

The Drug sequence encoder takes the sequence of the drugs taken by the patient into consideration. As opposed to modeling a single drug and its ADR, we model the entire drug sequence since multiple drugs taken in a given sequence or together can trigger certain ADRs compared to when a single drug is considered. We treat the drug sequence as a text and each drug name as tokens, similar to what is done in natural language processing. Therefore, it is important to choose an encoder instead of model the sequence and take the context around a given drug name into consideration. There are two common encoders for this task, the Bidirectional Long Short-Term Memory (BiLSTM) and the Transformers. Two variations of the BiLSTM used, word-based and character-based BiLSTM. While the performance of both models was comparable as we will see in the results, we believe that the character-based one generalizes better on drugs whose names are absent or less frequent in the dataset, making the model more robust. The resulting sequence is then tokenized using learned vocabulary and then passed to a word/character embedding layer where representations for each token is learned. Embeddings then are passed to a single-layer BiLSTM [20] to encode the sequence. Finally, the representation of the last hidden layer of the BiLSTM encoder is used as the sequence representation. Outputs of each time step are viewed as the encodings of each drug in the

sequence. Figure 3 shows a schematic of the word-based LSTM encoder inputs.

Transformer encoder models showed great performance in sequence modeling, especially on text-based sequences, due to their capability of producing contextual representations through self-attention [21]. We experimented with Transformer encoders to encode the drug sequences. We experimented with two famous architectures; small architecture using a transformer with 4 layers and 8 attention heads in each layer (L4H8), and large transformer architecture with 12 layers and 64 attention heads in each layer, denoted as (L12H64).

4.5 Adverse Event Reaction Encode

ADRs are written in plain English, their form is non-standardized, and they are more descriptive (e.g., high blood pressure, patient suffers from high blood pressure). Therefore, pre-trained language models are more suitable encoders for this type of input. Furthermore, modern language models, RoBERTa [22] included, use BPE in their tokenizers, which makes the model more robust towards variations in writing styles, grammatical errors, or misspellings. In this work, we utilize Huggingface's pre-trained roberta-base model and we unfreeze the top two layers only to be fine-tuned on our target dataset distribution.

4.6 Projection Layers

The output of the last encoder state from the drug sequence encoder, the CLS representation case of Transformers and the concatenation of the last forward and backward representation of the BiLSTM ($h^{\rightarrow}||h^{\leftarrow}$) are concatenated with the CLS token representation of the ADR RoBERTa encoder. The concatenated vector is then passed through two fully-connected layers and then softmax to output the likelihood of the drug-reaction pair. Training loss is then computed using binary cross entropy (BCE).

5. EXPERIMENTS AND RESULTS

5.1 Dataset

We split the cases in the FARES dataset into training (70%), validation (10%) and test (20%) datasets. A detailed summary of the number of cases in each dataset is presented in Table 1 and overall statistics of the drug sequences and reactions are presented on Table 2.

5.2 Model Setup

As mentioned earlier, for the drug sequence encoder, we experimented with biLSTM and Transformers, while for the ADR, we used the RoBERTa encoder. All models are randomly initialized using Xavier initialization. A higher dropout rate is used in the larger model to avoid overfitting.

We experimented with two architectures for the drug sequence transformer encoder; small architecture using a transformer with 4 layers and 8 attention heads in each layer (L4H8), and large transformer architecture with 12 layers and 64 attention heads in each layer, denoted as (L12H64) [20].

5.3 Results

As mentioned earlier, we experimented with multiple drug-sequence encoders: 1) word-based BiLSTM, 2) character-based LSTM, 3) transformer (L4H8), and 4) transformer (L12H64). We notice, and without much surprise, that the performance of transformer encoders, L4H8 and L12H64, surpassed the performance of the BiLSTM model by 7% and 10%, respectively. Transformers are able to allocate varying degrees of significance to distinct segments of the input sequence due to their self-attention mechanism. Transformers are able to extract more intricate patterns and relationships from the data than LSTMs because of their versatility. We then performed quantitative and qualitative evaluation to verify the correctness of our approach. Transformers are able to allocate varying degrees of significance to distinct segments of the input sequence due to their self-attention mechanism. Transformers are able to extract more intricate patterns and relationships from the data than LSTMs because of their versatility. We first looked at the model predictions on whether the drug sequence and reaction are associated with each other or not, this is a binary classification problem. For that we computed the accuracy and F1-score as shown in Table 3. We can see that character-based LSTM outperformed the word-based LSTM by about 3% in terms of F1-score. The best performing model is the larger transformer (L12H64), adding close to 12.5% improvement in F1-score over the baseline (word-based BiLSTM). Those models were trained for 55k steps on a Nvidia T4 GPU.

Table 3. Performance Metrics

Model	Accuracy	F1
Word-based BiLSTM (baseline)	0.8	0.74
Character-based BiLSTM	0.84	0.77
Transformer L4H8	0.91	0.84
Transformer L12H64	0.94	0.87

In addition to the quantitative analysis above, we performed qualitative analysis to verify the model is in fact learning good drug-sequence representations. To do that, we randomly sampled 20k drug sequences covering four adverse events (breast cancer, dizziness, physical disability and hyperbilirubinemia), then we visualized their embeddings using t-SNE plot (Figure 4). As you can see in Figure 5, drug sequence embeddings for each reaction are grouped together, validating that sequences with similar reactions are closer in the embeddings space, while sequences with different reactions are distant from each other.

We looked further into four of the clusters (hyperbilirubinaemia, breast cancer, dizziness, and physical disability) shown in Figure 4. We selected three drug sequences from each of the three clusters and plotted the heatmap generated based on the cosine similarity as shown in Figure 4. We notice a pattern in the heatmap that again validates that the model is learning good drug sequence representation. In Figure 5, (x_1, x_2, x_3) are sampled from one cluster, (t_1, t_2, t_3) are sampled from a second cluster, (s_1, s_2, s_3) are sampled from a third cluster and (c_1, c_2, c_3) are sampled from the fourth cluster. We notice clear groupings among the selected sequences. For instance, (t_1, t_2, t_3) are very close to each other with cosine similarity greater than 0.82, while they are distant from all other sequences with cosine similarity to sequences in other clusters dropping to a range between 0.22-0.77.

For a reference, we include the drug sequences and their reactions presented in Figure 4 in Table 4.

Table 4. Drug Sequences From Figure 3 And Their Corresponding ADR. Drug Sequences Can Appear Multiple Times In The Dataset With Different Reactions Sequences, Enlisted Here Is A Random Sample Per Each.

	Drug Sequence	ADRs
c_1	GANCICLOVIR/ESMOLOL HYDROCHLORIDE/HYDRALAZINE/HYDROCHLORIDE/VALGANCICLOVIR/COTRIMOXAZOLE	Hyperbilirubinaemia, Hepatomegaly
c_2	Olanzapine /ANAFRANIL/Centrum maternal	Hyperbilirubinaemia neonatal, Respiratory distress,Feeding disorder
c_3	CAPECITABINE./CAPECITABINE./QUINACRINE/QUINACRINE	Hyperbilirubinaemia, Oedema peripheral, Rash
s_1	RANITIDINE HYDROCHLORIDE/RANITIDINE HYDROCHLORIDE/RANITIDINE /ZANTAC	Breast cancer stage (III/I)
s_2	/XYREM/LISINOPRIL/COENZYME Q10/LEVOTHYROXINE/ATENOLOL/NUVIGIL/VITAMIN D3	Breast cancer, carcinoma, Delayed sleep phase
s_3	REVLIMID/PREDNISONE	Neutropenia, Influenza, Breast cancer
t_1	Infliximab/INDOCID/MORPHINE/ NAPROXEN./VOLTAREN	Dizziness, Colitis ulcerative, Pyrexia
t_2	CARBOPLATIN/TECENTRIQ/AVASTIN/PACLITAXEL	Malaise, Dizziness
t_3	ULTOMIRIS/ULTOMIRIS	Infusion related reaction, Dizziness
x_1	RISPERDAL/TEMAZEPAM/CITALOPRAM/MIRTAZAPINE/SERTRALINE/CIPRALEX/TRAZODONE	Physical disability, Sedation, Somnolence
x_2	ENTRESTO/LOSARTAN	Fatigue, Physical disability, Malaise, Dyspnoea
x_3	XELJANZ XR/XELJANZ XR/XELJANZ XR	Exercise tolerance decreased, Stress, Abdominal discomfort, Physical Disability

6. CONCLUSION

Our model aims at providing contextual distributed representation for the drug sequence by leveraging the drug sequence-reaction data pairs. The representations are learned by predicting whether a target adverse reaction can be associated

with drug sequence or not. Our evaluations have demonstrated the correctness of the learnt embeddings with an F1-score reaching 0.87 when using Transformer encoder for the drug sequences. The qualitative analysis has also shown how drug sequences are close to each other when sharing similar ADRs and distant otherwise.

7. LIMITATION AND FUTURE WORK

In this work, we reviewed distributed representation of drugs based on the drug alone. We also processed the drug sequence by removing duplicate drug names that were repeated multiple times in the sequence, which is not the ideal solution. FARES data contains data that differentiates the repeated drugs in a sequence. This data includes medication dosage, strength, route, form, duration that can be incorporated into the model to improve its performance. Additionally, the FARES data includes unstructured data that explains the reported case in free text, which can also be used to help predict the reactions. However, the unstructured data is not easily accessible.

In future work, we will reframe the problem as reaction prediction, rather than predicting the association between drug sequence and the reaction. This requires different model architecture that can also incorporate the additional data elements mentioned above.

Finally, it is important to note that the accuracy of the drug sequence embeddings is limited by the quality of the FARES data. In future work, we will consider validating a small sample of the data to verify the validity and accuracy of the reported adverse events.

REFERENCES:

[1] Edwards, I. Ralph, and Jeffrey K. Aronson. "Adverse drug reactions: definitions, diagnosis, and management." The lancet 356.9237 (2000): 1255-1259.

[2] .Ernst, Frank R., and Amy J. Grizzle. "Drug-related morbidity and mortality: updating the cost-of-illness model." Journal of the American Pharmaceutical Association (1996) 41.2 (2001): 192-199.

[3] Assaf, Rasha, Rashid Jayousi, and Amjad Rattrout. "Current State of Machine Learning Based Methods for Adverse Events Prediction." 2021 International Conference on Promising Electronic Technologies (ICPET). IEEE, 2021.

[4] https://www.fda.gov/

[5] Onakpoya, Igho J., Carl J. Heneghan, and Jeffrey K. Aronson. "Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: a systematic review of the world literature." BMC medicine 14.1 (2016): 1-11.

[6] Embi, Peter Joseph, et al. "Responding rapidly to FDA drug withdrawals." Journal of Medical Internet Research 8.3 (2006).

[7] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature 521.7553 (2015): 436-444.

[8] Urban, Gregor, et al. "Deep learning for drug discovery and cancer research: Automated analysis of vascularization images." IEEE/ACM transactions on computational biology and bioinformatics 16.3 (2018): 1029-1035.

[9] Vilar, Santiago, et al. "Drug—drug interaction through molecular structure similarity analysis." Journal of the American Medical Informatics Association 19.6 (2012): 1066-1074.

[10] Portanova, Jake, et al. "aer2vec: distributed representations of adverse event reporting system data as a means to identify drug/side-effect associations." AMIA Annual Symposium Proceedings. Vol. 2019. American Medical Informatics Association, 2019.

[11] Lee, Chun Yen, and Yi-Ping Phoebe Chen. "Prediction of drug adverse events using deep learning in pharmaceutical discovery." Briefings in Bioinformatics 22.2 (2021): 1884-1901.

[12] Dey, Sanjoy, et al. "Predicting adverse drug reactions through interpretable deep learning framework." BMC bioinformatics 19.21 (2018): 1-13. .

[13] Kuhn, Michael, et al. "The SIDER database of drugs and side effects." Nucleic acids research 44.D1 (2016): D1075-D1079.

[14] Wu, Zengrui, et al. "Network-based methods for prediction of drug-target interactions." Frontiers in pharmacology 9 (2018): 1134.

[15] Deac, Andreea, et al. "Drug-drug adverse effect prediction with graph co-attention." arXiv preprint arXiv:1905.00534 (2019).

[16] Ménard, Timothé, et al. "Enabling data-driven clinical quality assurance: predicting adverse event reporting in clinical trials using

machine learning." Drug safety 42 (2019): 1045-1053.
[17] Anastopoulos, Ioannis N., et al. "Multi-Drug Featurization and Deep Learning Improve Patient-Specific Predictions of Adverse Events." International Journal of Environmental Research and Public Health 18.5 (2021): 2600.
[18] Khaleel, Mohammad Ali, et al. "A standardized dataset of a spontaneous adverse event reporting system." *Healthcare*. Vol. 10. No. 3. MDPI, 2022.
[19] Robinson, Joshua, et al. "Contrastive learning with hard negative samples." arXiv preprint arXiv:2010.04592 (2020).
[20] Schuster, Mike, and Kuldip K. Paliwal. "Bidirectional recurrent neural networks." IEEE transactions on Signal Processing 45.11 (1997): 2673-2681.
[21] Shaw, Peter, Jakob Uszkoreit, and Ashish Vaswani. "Self-attention with relative position representations." arXiv preprint arXiv:1803.02155 (2018)..
[22] Liu, Yinhan, et al. "Roberta: A robustly optimized bert pretraining approach." arXiv preprint arXiv:1907.11692 (2019).

APPENDIXES:

Journal of Theoretical and Applied Information Technology
29th February 2024. Vol.102. No 4

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

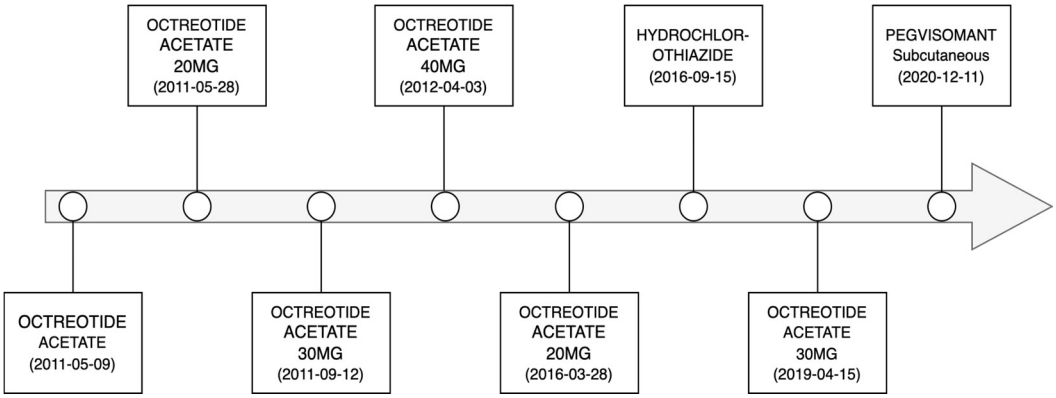

Figure 1. Example Drug Sequence From FARES Data

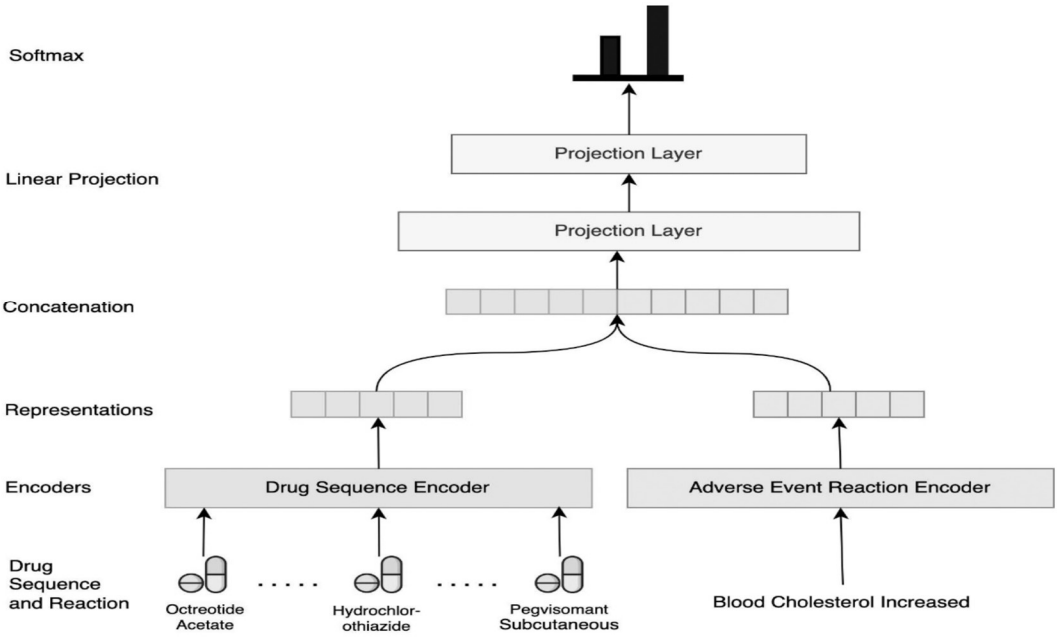

Figure 2. Model Architecture

Journal of Theoretical and Applied Information Technology
29th February 2024. Vol.102. No 4

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

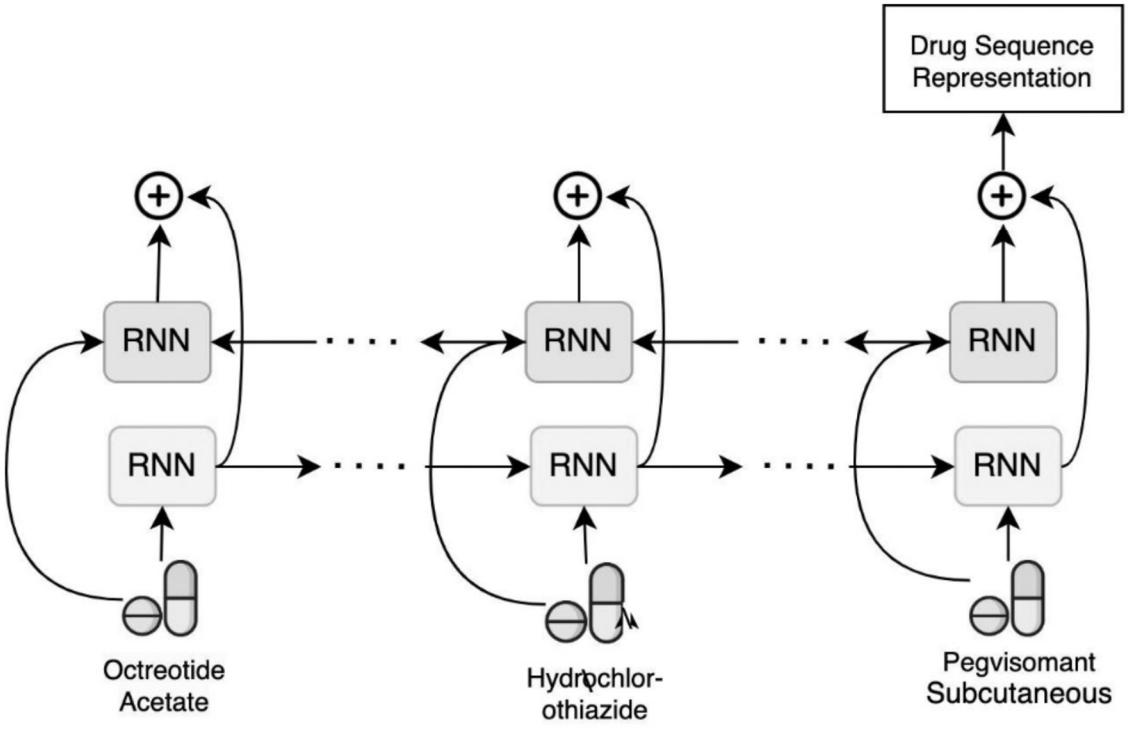

Figure 3. Bilstm Drug Sequence Encoder

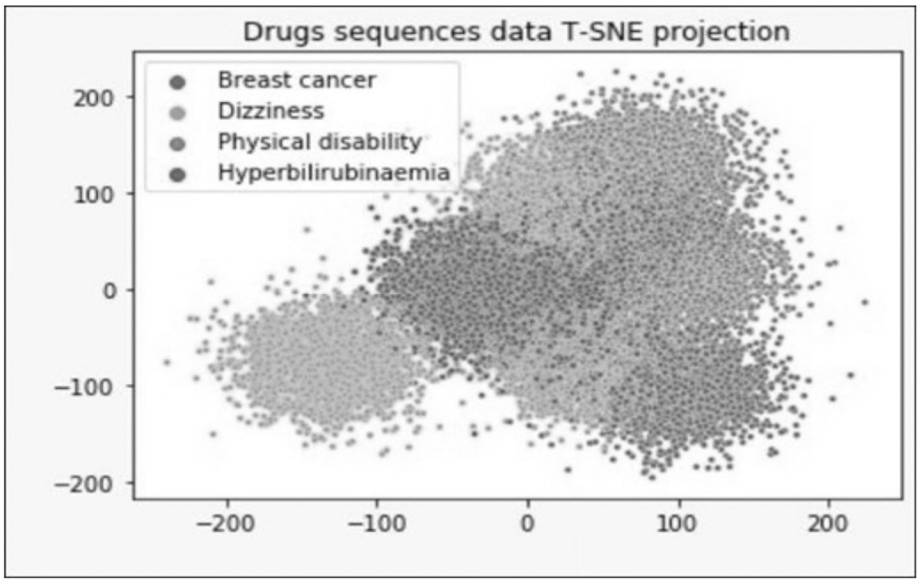

Figure 4: T-SNE Visualization Of 4 Clusters Of Drug Sequences Sampled By Their Respective Reactions.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

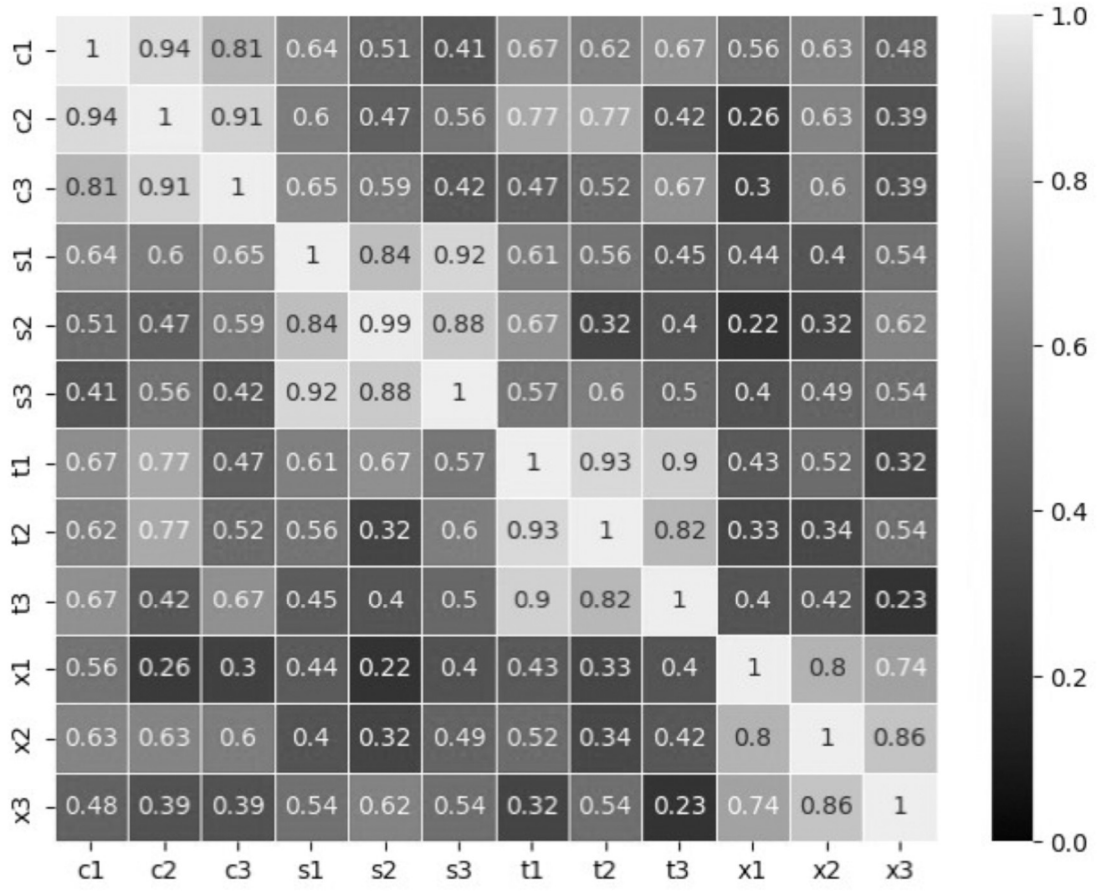

Figure 5: Cosine Similarity Matrix Between 3 Random Samples From Each Cluster In Figure 4

Table 1. Drug Sequence Length And Reactions For The Training And Validation Datasets

	Training Set			Validation Set		
	Min.	Avg.	Max.	Min.	Avg.	Max.
Sequence length (folded)	1	17	21	1	15	23
Sequence length (unfolded)	1	11	18	1	9	16
Number of tokens per reactions	1	7	20	1	15	17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2. Training, Validation And Test Set Sizes Before And After Folding The Adrs.

	Before ADR folding	After unfolding
Training set size	600K	1.87M
Validation set size	160K	490K
Test set size	40K	135K