Artificial Neural Network and Latent Semantic Analysis for Adverse Drug Reaction Detection

Ahmed Adil Nafea (iD)   *Nazlia Omar* (iD)   *Zohaa Mubarak Al-qfail* (iD)

Center for Artificial Intelligence Technology (CAIT), Faculty of Information Science and Technology Universiti Kebangsaan Malaysia (UKM), Bangi, Selangor, Malaysia
*Corresponding authors: ahmed.adil.nafea@gmail.com
E-mail addresses: nazlia@ukm.edu.my, alqfailzohaa85@gmail.com

Abstract:

Adverse drug reactions (ADR) are important information for verifying the view of the patient on a particular drug. Regular user comments and reviews have been considered during the data collection process to extract ADR mentions, when the user reported a side effect after taking a specific medication. In the literature, most researchers focused on machine learning techniques to detect ADR. These methods train the classification model using annotated medical review data. Yet, there are still many challenging issues that face ADR extraction, especially the accuracy of detection. The main aim of this study is to propose LSA with ANN classifiers for ADR detection. The findings show the effectiveness of utilizing LSA with ANN in extracting ADR.

Keywords: Adverse Drug Reaction, Artificial Neural Network, Classification, Deep Learning, Latent Semantic Analysis.

Introduction:

Adverse drug reaction (ADR) is defined as a substantially injurious or unpleasant reaction as a consequence of an intervention relating to a medicinal product utilized, anticipating potential administering hazards and warranting avoidance or immediate care, or change of dosage scheme or removal of a product [1,2]. ADR represents severe problems throughout the world. These problems can complicate a patient's medical conditions or increase morbidity, including death. A previous study showed around 100,000 deaths from medical errors in the US in 2000, while 7,000 were attributable to medication responses [3]. Recently, a new type of product review has caught researchers' attention which is the medical review [4]. In this regard, users can evaluate drug products by describing their experience when taking such medicine [5]. Hence, several side effects and other related-medical entities would be mentioned. Therefore, a task has emerged to detect these mentions which are known as Adverse Drug Reaction (ADR) detection [6, 7].

In the literature, ADR detection aims to automatically determine whether a sentence is related to an ADR[8]. Most studies have crawled data from social media sites like Twitter or websites for drugs or medication reactions. Regular user comments and reviews were considered during the data collection process to extract ADR mentions. For instance, an ADR of "dizzy" appears in a study that says, "After taking this drug, I felt dizzy," where the user reports an adverse effect they experienced after taking a particular drug.

Most researchers were focusing on machine learning techniques to detect ADR [9,10]. Such strategies used medical review annotated data to train a classification model [11,12]. Yet, there are still many challenging issues that face this automatic ADR extraction [13]. One of these issues is the accuracy of detection. This could be due to the fact that most of the corpora used for ADR detection are on a small scale [14,15]. Applying machine learning approaches usually demands large data sizes for better performance. However, incorporating additional corpora sampled from different sources for the training may introduce noises and impact the performance of the machine learning or neural networks [14,16]. Methods like Latent Semantic

Analysis (LSA) which exploits the contextual-usage meaning of words by statistical computations could assist in improving the performance of ADR using a small dataset [17].

The main objective of this paper is to propose LSA with an ANN classifier for the purpose of ADR detection. The findings of this research may be included in the medicinal opinion mining approach to assist not only patients in considering the medicine before taking it but also physicians and organizations associated with drug manufacturers in taking user input into account when making decisions.

The paper is constructed as follows: In section II, we discuss the related works. In section III, we present our proposed method. After that, we explain the experimental results and discussion in section IV. Then we complete our findings in section V with future recommendations and outcomes.

Related Work

The literature indicates that many researchers are gaining interest in the task of ADR detection [10-13, 18]. The initial benchmark dataset for medical evaluations was provided by [19]. In order to distinguish reviews from ADR, the authors used trigger phrases in addition to the rule-based approach. Additionally, this study suggested classifying ADR not reported to the US Food and Drug Administration (FDA) by automatically collecting ADR from user comments on a variety of social media sites.

Pain et al. [11] presented an ADR detection technique using the Support Vector Machine (SVM) classifier. The proposed solution made use of a list of hashtags and keywords that commonly appeared in ADR. The researchers developed an automated drug-effect identification system using medical review data obtained from Twitter. The suggested traits can recognize a wide range of drug-effect entities. Their study covered the creation of post-marketing surveillance (PMS) techniques tailored to Twitter's text.

Ebrahimi et al. [5] presented a technique of ADR extraction using SVM. Similarly, the authors have utilized the trigger terms with more medical concepts. To ascertain the side effects of medications from medical reviews, they used a set of medical concepts with properly identified entities as trigger phrases. The syntactic tag of phrases was identified via POS tagging. To find pharmacological side effects, a rule-based classification technique and SVM were used. As a subtask to identifying implicit views in medical literature and differentiating side effects and disease symptoms, this research devised a method to identify side effects in pharmaceutical reports.

Plachouras et al. [12] have used N-gram representation and a set of gazetteer features to adverse drug events extraction from Twitter reviews. The strategy for supporting widespread pharmacovigilance was presented in this study. Through the training and testing of a supervised binary classifier, they addressed the issue of extracting negative events from tweets. In order to accommodate the final extraction, the author utilized POS tags, sentiment analysis, a collection of gazetteers, words, and keywords, and surface features. These techniques were combined with the SVM classification method[20]. The primary drawback of this study is its reliance on trigger phrases, even when the semantic component might be ignored.

Kiritchenko et al. [9] proposed two collaborative tasks during the AMIA-2017 Workshop on Social Media Mining for Health Applications (SMM4H). The first task entailed categorizing tweets that mentioned ADR, whereas the second one involved categorizing tweets that discussed individual pharmaceutical usage. Using an SVM approach for ADR extraction, vector machine classifiers were trained with a range of surface-specific features, domain-specific features, and feelings for both tasks. To improve detection accuracy, the researchers filtered the trigger terms and used a domain-specific one. Medical reviews on Twitter were used in experiments.

Emadzadeh et al. [18] proposed to enhance the performance of extracting ADR by combining the Unified Medical Language System (UMLS) with latent semantic analysis and hybrid semantic analysis. This research presents a modular NLP pipeline for standardizing the mapping between ADRs' common names and their standardized Identifiers. They employed a publicly accessible, annotated corpus for their investigation, and they were successful in achieving an F-measure of 0.624.

Yousuf et al. [10] analyzed how to extract adverse drug reactions (ADR) from social networks where users discuss specific medications. The majority of obtaining entities rely on certain phrases, sometimes referred to as trigger terms, which may exist before or after ADR. However, such definitions should be updated, amended, and regularly revised. The goal of this study was to propose an expansion of the trigger phrase based on several N-gram representations. Two document representations were used: TF-IDF and TF. The tests have an F-measure of 0.69 and used secondary information from medication websites.

Nafea Ahmed et al. [13] aim to provide a semantic approach based on machine learning algorithms with latent semantic analysis (LSA) to enhance the identification of ADR. The studies utilized a benchmark dataset, several pre-processing techniques, and three classifiers trained on the proposed LSA, namely Naive Bayes, Support Vector Machine, and linear regression. Two document representations, TF and TFIDF, were utilized. The study demonstrated good performance in terms of extracting ADR with an F- measure of 82%. However, we believe the accuracy of detection can still be improved. Hence, this study aims to propose an LSA with a deep learning technique using ANN for improving the detection of ADR. Applying LSA to the architecture of deep

learning could improve the performance of ADR detection accuracy.

Research Methodology

The methodology involves five stages, and this is shown in Fig. 1. Firstly, annotated drug reviews are prepared using a dataset from Yates et al. [19] benchmark dataset that has had some structure updated by Yousef et al. [10] by adding more useful data fields. Secondly, preprocessing tasks were performed like stop word removal, tokenization and stemming. Thirdly, the terms are represented in vector space using TF and TF-IDF. Fourthly, semantic analysis is performed using LSA. Finally, the classification is done by using ANN. The following describes the methodology in more detail.
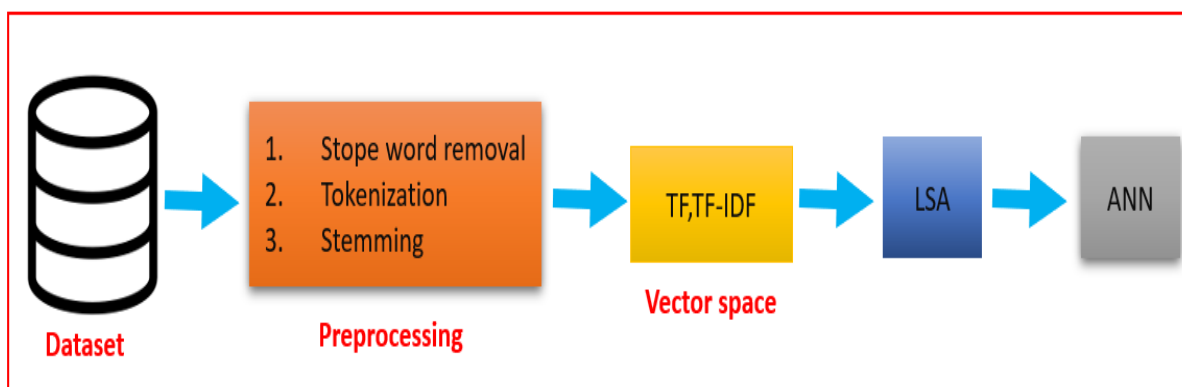

Figure 1. Methodology.

Dataset

This proposed approach utilized the same benchmark dataset by Yates et al. [19], with some structure updated from Yousef et al. [10] by adding more insightful columns of the data. There are 2500 reviews in the dataset utilized for this study (with 246 labelled documents). One or more sentences can be found in every document. There are 945 sentences in the texts extracted from Twitter. There are 982 ADRs in total among all texts. The language used to write these documents is English. The review data is gathered from three popular social media drug review sites, namely, askapatient.com, drugratingz.com, and drugs.com.

Preprocessing

In this stage, the text will undergo several preprocessing steps. The description is as follows:

- **Stop word removal** seeks to remove the common words that do not contain significant information. These terms are often deleted during preprocessing to decrease the amount of noise data or features that are not very useful [21,22]. Typically, these are terms that are used frequently in the text, such as "the" "a" "an" and "of".

- **Tokenization** The purpose of tokenization is to convert the text into a set of sentences, and then those words are turned into a set of tokens (words) [23].

- **Stemming** In this work Porter's Stemmer method was applied [24]. It is one of the most well-known stemming techniques since 1980. It is based on the idea that suffixes are made up of a range of smaller and simpler suffixes in the English language.

Term Representation

In this step, the data will represent the word frequency in the documents by the TF or TF-IDF. Term Frequency (TF) The term frequency is used in this process to indicate how many times the word appears in document [25]. The following Eq.1 is used to address the frequency problems:

$$W_d\,(t) = \ TD(t, d\,) \qquad\qquad 1$$

the word is t and the frequency document is d.

Inverse Document Frequency (IDF) seeks to have high weight for uncommon conditions, and low typical conditions weights [26]. The Eq.2 is as follows:

$$IDF_t = \log n\left(\frac{N}{N_t}\right) \quad (t, d\,) \qquad\qquad 2$$

where N_t is the number of documents that include the word and N denotes the total number of documents.

Term Frequency with Inverse Document Frequency TF-IDF This approach combines two separate TF and IDF approaches [25,27]. The following is the weighting Eq.3:

$$W_t = TF(t, d). IDF_t \qquad 3$$

where TF means a term frequency and IDF means the inverse document frequency of term.

Latent Semantic Analysis

In NLP processing, LSA is frequently used to determine the similarities between two different text classes [28]. LSA aims to investigate the links between the meanings of the words, expressions, and concepts in both sets of documents by creating a vector space, where the documents are represented in the columns and the terms are displayed in the rows. To determine the semantics, LSA first employs either TF or TF-IDF, where all the unique words are gathered in separate attributes. Therefore, LSA takes either a CV or TF-IDF matrix as an input and outputs a dimensioned matrix with more values that accurately reflect the semantics of each word. This is achieved using a method known as Singular Value Decomposition (SVD). Minimizing the number of rows while maintaining the structure of similarity between columns is one goal of SVD, as shown in Fig. 2. The SVD computation is shown in the following Eq.4.
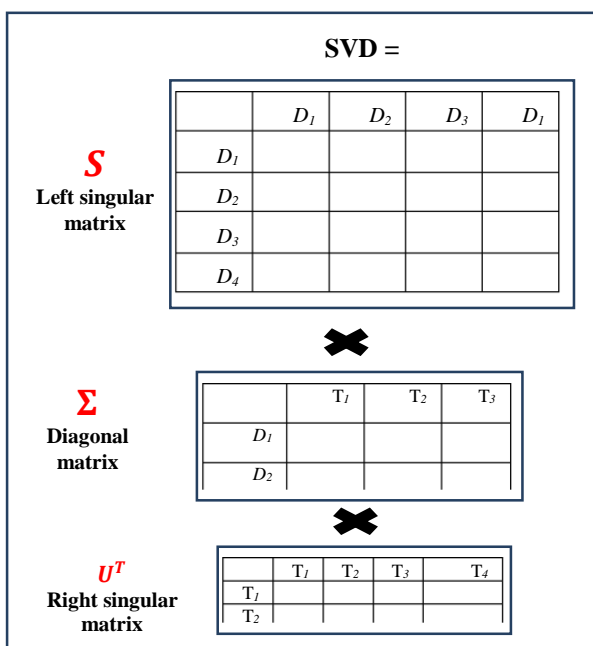
$$SVD = S\Sigma U^T \qquad 4$$

Figure 2. Matrices of SVD [13]**.**

Deep learning approach (ANN)

At this stage, the complicated matrix has completed obtaining the semantic information. TF or TF-IDF is used by ANN, where all the distinctive words are sorted in disconnected attributes. Therefore, a CV or TF-IDF matrix can be used as the input for LSA and produces the same dimension matrix but with more advanced values that accurately reflect the semantics of each word. This is accomplished using the SVD. Then it will be classified using the proposed method ANN algorithm.

The data is divided into 30% for testing and 70% for training is the same within the baseline. The learning rate is 0.01, the type of activation function is ReLU, the number of hidden layers is 2 the first hidden layer contains 64 units and the second one contains 32 units, the random state is 1 and the maximum iteration is 300.

The main structure of an artificial neural network (ANN) as shown in Fig. 3. This is the main building block that the proposed approach is based. A neural network consists of 3 types of layers [29]:

- **Input layer**: The first layer is responsible for transferring input data to the next layer.
- **Hidden layers:** are responsible for understanding the connection between input and output values.
- **Output layer**: This layer produces the ANN's prediction result based on the given inputs.

The first process in the input layer starts with the input data and any number of random weights. Then feedforward process follows, producing the prediction values as an output result. After that, the prediction and real value are compared to evaluate the loss score. This calculation is done with the loss function. Then, the backpropagation is the next station.

Backpropagation merely updates the weights for each weight in the ANN by computing the derivative of the loss score. These derivatives are known as gradients. This gradient-based optimization is the mechanism of the ANN. Using backpropagation, the loss score can be reduced via an update of the weights. Ideally, the network should figure out the best weights to use to predict output with the least amount of error.
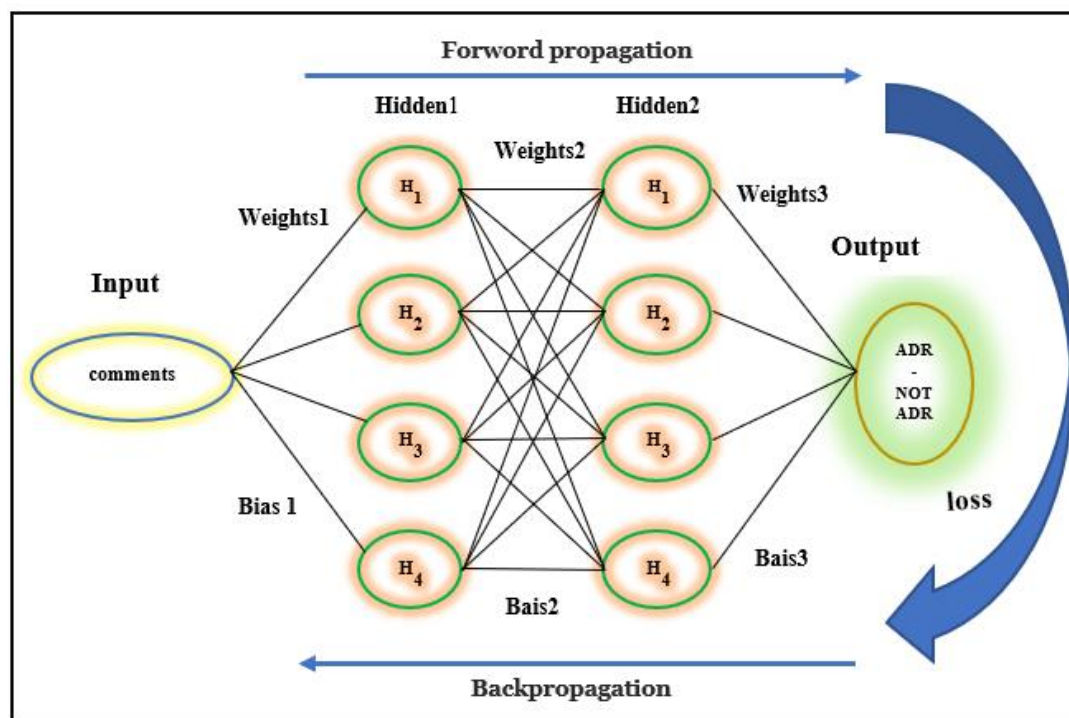
Figure 3. The proposed model ANN.

Evaluation

After applying the classification ADR using ANN, it is crucial to assess the outcomes of the classification achieved by ANN. To accomplish that, it is essential to underline some significant factors that indicate the correctness of the classification. For this purpose, it can be represented using the correct classified ADR known as TP, whereas FP is the incorrect classified ADR. Moreover, FN is the wrongly refused classified ADR, and lastly, TN rightly refused classified ADR. Therefore, the precision, recall and f-measure can be determined using the following Eq.5, Eq.6 and Eq.7 respectively [30,27].

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad 5$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{TN}} \qquad 6$$

$$\begin{aligned}\text{F} &- \text{measure} \\ &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad 7\end{aligned}$$

Result and Discussion:

This section shows the outcomes of the suggested combination of LSA and ANN models. The experimental setting aims to compare the performance of the proposed research with the baseline [13]. The baseline used the same data set from a benchmark dataset. The annotated ADR review dataset used is by Yates et al. [19] which Yousef et al. [10] updated by adding additional

meaningful columns to the data. The baseline utilized the same suggested technique through the feature representation, which employed LSA with TF or TF-IDF. Then, testing and training were conducted using the same distribution as the baseline. The data is divided into 30% for testing and 70% for training. The baseline employed different machine learning methods classifiers like SVM, NB, and LR whereas this study used ANN. The outcomes of the proposed ANN via LSA through the TF and TF-IDF with baseline are shown in Table. 1.

The classification results based on F-measure are presented in Fig. 4, showing the comparison between the ADR baseline research using machine learning [13] opposite the proposed work using ANN. The findings demonstrated that the presented technique employing ANN showed improvement in terms of f-measure when compared with the baseline research which utilized other machine learning algorithms. The proposed ANN using the TF and ANN showed enhanced f-measure results, achieving 85% as opposed to the baseline LR of 82%. The proposed ANN using the TF-IDF showed enhanced f-measure results, achieving 83% as opposed to the baseline LR 80%. The superiority of using LSA with ANN is due to the fact that semantic correspondences have been identified correctly, as opposed to using LSA with machine learning algorithms. In terms of detecting ADRs, the suggested LSA with ANN performs better than the baseline. This result suggests that LSA with ANN approach to extracting ADRs is promising.

Table 1. A comparison of results on the proposed approach and baseline

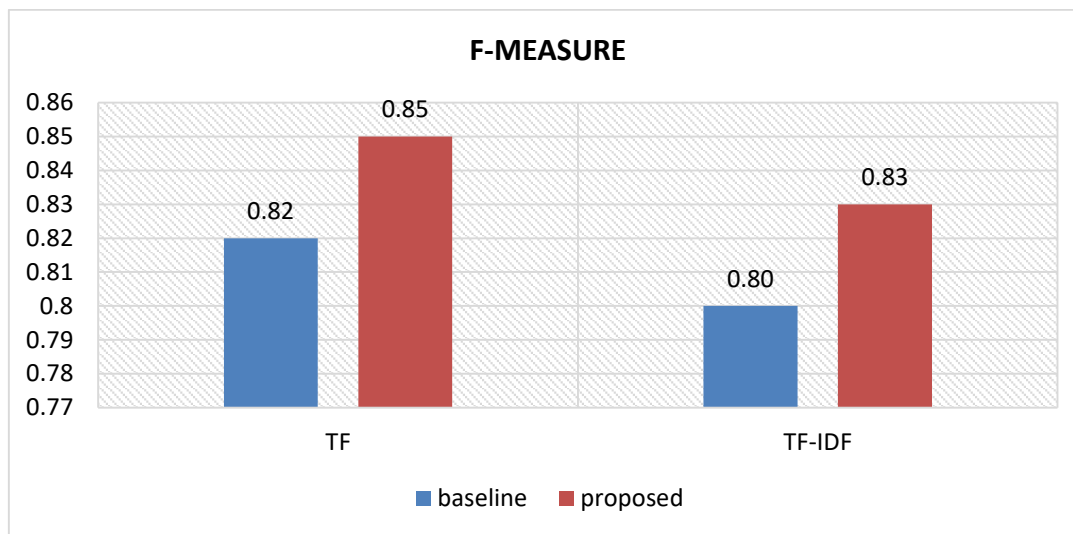|	TF			TF-IDF		
	Precision	**Recall**	**F-measure**	**Precision**	**Recall**	**F-measure**
Baseline	0.83	0.82	**<u>0.82</u>**	0.82	0.81	**<u>0.80</u>**
proposed	0.85	0.85	**<u>00.85</u>**	0.83	0.83	**<u>0.83</u>**

Figure 4. A Comparison between baseline and proposed results.

In addition to the traditional baseline, which made use of standard strategies like a machine learning approach, it is important to discuss cutting-edge strategies that used deep learning techniques. Zhang et al. [2] used Twitter ADR data and this proposed adversarial transfer learning architecture for the ADR challenge performs with an F1 of 68.58%. Yousef et al. [10] have created trigger terms to identify ADR using machine learning categorization techniques. They succeeded in achieving an f-measure of 69%. Lee et al. [31] have extracted ADRs using CNN's deep learning method, achieving an f-measure of 64.5%. Cocos et al. [32] have achieved an f-measure of 75.5% using a deep learning method of RNN to extract ADRs. Wang et al. [33] which made use of deep neural network techniques have achieved an f-measure of 84.4%. Due to the differing dataset, it is not feasible to compare these findings to the suggested approach. The performance of these deep learning techniques depends on many factors and the size of the dataset could be one of them. The proposed ANN is still seen to be competitive especially when the data used is on a small scale.

Conclusion:

The results showed that the proposed approach outperforms the baseline by achieving 85 % of the f-measure via TF with ANN compared to the f-measure obtained by the baseline of 82%. Therefore, the proposed method demonstrated that the use of LSA with ANN had improved ADR detection. This finding implies the effectiveness of using LSA with ANN in extracting ADRs. The limitation of this study exploring real-time reviews would contribute towards discovering new side-effects of new drugs such as covid 19 drugs. in the future could apply sophisticated word embedding with the architecture of deep learning. It could improve the performance of ADR detection.

Authors' declaration:
- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are mine ours. Besides, the Figures and images, which are not mine ours, have been given the permission for re-publication attached with the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee in UKM.

Ethics Approval
The material in this article is original and has not been published in any journal previously. The lead author confirms that all co-authors have reviewed and approved the work and that there are no ethical issues.

Authors' contributions statement:
A.A.N, N.O, Z.M.A contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript.

References:

1. Zhang T, Lin H, Xu B, Yang L, Wang J, Duan X. Adversarial neural network with sentiment-aware attention for detecting adverse drug reactions. J Biomed Inform. 2021; 123: 103896. https://dx.doi.org/10.1016/j.jbi.2021.103896

2. Zhang T, Lin H, Ren Y, Yang Z, Wang J, Duan X, et al. Identifying adverse drug reaction entities from social media with adversarial transfer learning model. Neurocomputing. 2021; 453: 254–62. https://dx.doi.org/10.1016/j.neucom.2021.05.007

3. Ray SD, Luzum JA, Gray JP, Stohs SJ. Focus on pharmacogenomics, phytonutrient-drug interactions and COVID-19 vaccines: Perspectives on ADRs, ADEs, and SEDs. Side Eff Drugs Annu. 2021; 43:xxv. https://dx.doi.org/10.1016/S0378-6080(21)00054-4

4. Li R, Curtis K, Zaidi ST, Van C, Castelino R. A new paradigm in adverse drug reaction reporting: consolidating the evidence for an intervention to improve reporting. Expert Opin Drug Saf. 2022; 21(9): 1193–204. https://dx.doi.org/10.1080/14740338.2022.2118712

5. Ebrahimi M, Yazdavar AH, Salim N, Eltyeb S. Recognition of side effects as implicit-opinion words in drug reviews. Online Inf Rev. 2016; 40(7): 1018–32. https://dx.doi.org/10.1108/OIR-06-2015-0208

6. Kumar NH, Narendra B, Upendra K, Rajesh K. A review on Adverse drug reactions monitoring and reporting. Int J Pharm Res Technol. 2019; 9(2): 12–5. https://dx.doi.org/10.31838/ijprt/10.01.09

7. Yousef RNM, Tiun S, Omar N, Alshari EM. Enhance medical sentiment vectors through document embedding using recurrent neural network. Int J Adv Comput Sci Appl. 2020; 11(4): 372-378. https://dx.doi.org/10.14569/IJACSA.2020.0110452

8. Siuly S, Lee I, Huang Z, Zhou R, Wang H, Xiang W. Health Information Science. 1st ed. Springer International Publishing; 2018. https://dx.doi.org/10.1007/978-3-030-01078-2

9. Kiritchenko S, Mohammad SM, Morin J, de Bruijn B. NRC-Canada at SMM4H shared task: classifying Tweets mentioning adverse drug reactions and medication intake. arXiv Prepr arXiv180504558. 2018. https://dx.doi.org//10.48550/arXiv.1805.04558

10. Yousef RN, Tiun S, Omar N. Extended Trigger Terms for Extracting Adverse Drug Reactions in Social Media Texts. J Comput Sci. 2019; 15(6): 873–9. https://dx.doi.org/10.3844/jcssp.2019.873.879

11. Pain J, Levacher J, Quinquenel A, Belz A. Analysis of Twitter Data for Postmarketing Surveillance in Pharmacovigilanc. In: Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT). COLING 2016 Organ Commit. 2016 : 94–101. https://aclanthology.org/W16-3914

12. Plachouras V, Leidner JL, Garrow AG. Quantifying self-reported adverse drug events on Twitter: signal and topic analysis. Proc 7th Int Conf Soc Media Soc. ACM Press. 2016: 1–10. https://dx.doi.org/10.1145/2930971.2930977

13. Nafea AA, Omar N, AL-Ani MM. Adverse Drug Reaction Detection Using Latent Semantic Analysis. J Comput Sci. 2021; 17(10): 960–70. https://dx.doi.org/10.3844/jcssp.2021.960.970

14. Li Z, Yang Z, Luo L, Xiang Y, Lin H. Exploiting adversarial transfer learning for adverse drug reaction detection from texts. J Biomed Inform. 2020; 106: 103431. https://dx.doi.org/10.1016/j.jbi.2020.103431

15. Al-Shareeda MA, Manickam S, Laghari SA, Jaisan A. Replay-Attack Detection and Prevention Mechanism in Industry 4.0 Landscape for Secure SECS/GEM Communications. Sustainability. 2022; 14(23): 15900. https://dx.doi.org/10.3390/su142315900

16. Al-Shareeda MA, Manickam S. COVID-19 Vehicle Based on an Efficient Mutual Authentication Scheme for 5G-Enabled Vehicular Fog Computing. Int J Environ Res Public Health. 2022; 19(23): 15618. https://dx.doi.org/10.3390/ijerph192315618

17. Al-Sabahi K, Zhang Z, Long J, Alwesabi K. An enhanced latent semantic analysis approach for arabic document summarization. Arab J Sci Eng. 2018; 43(12): 8079–94. https://dx.doi.org/10.1007/s13369-018-3286-z

18. Emadzadeh E, Sarker A, Nikfarjam A, Gonzalez G. Hybrid semantic analysis for mapping adverse drug reaction mentions in tweets to medical terminology. AMIA Annu Symp Proc. PubMed Central. 2018: 679–688. PMCID: PMC5977584

19. Yates A, Goharian N. ADRTrace: detecting expected and unexpected adverse drug reactions from user reviews on social media sites. IAdvances in Information Retrieval: 35th European Conference on IR Research. 2013, Moscow, Russia, March 24-27, 2013 Proceedings 35. Springer Berlin Heidelberg; 2013: 816–9. https://dx.doi.org/10.1007/978-3-642-36973-5_92

20. Abdulmajeed AA, Tawfeeq TM, Al-jawaherry MA. Constructing a Software Tool for Detecting Face Mask-wearing by Machine Learning. Baghdad Sci J. 2022; 19(3): 642. https://dx.doi.org/10.21123/bsj.2022.19.3.0642

21. Kaur J, Buttar PK. A systematic review on stopword removal algorithms. Int J Futur Revolut Comput Sci Commun Eng. 2018; 4(4): 207–10. https://www.ijfrcsce.org/index.php/ijfrcsce/article/view/1499

22. Oliinyk VA, Vysotska V, Burov Y, Mykich K, Fernandes V. Propaganda detection in text data based on NLP and machine learning. In: CEUR Workshop Proceedings. MoMLeT+DS. 2020: 132–44. https://ceur-ws.org/Vol-2631/paper10.pdf

23. Chary M, Parikh S, Manini AF, Boyer EW, Radeos M. A review of natural language processing in medical education. West J Emerg Med. 2019; 20(1): 78. https://dx.doi.org/10.5811/westjem.2018.11.39725

24. Patel R, Passi K. Sentiment Analysis on Twitter Data of World Cup Soccer Tournament Using Machine Learning. IoT. 2020; 1(2): 218–39. https://dx.doi.org/10.3390/iot1020014

25. Mohammed M, Omar N. Question classification based on Bloom's taxonomy cognitive domain using modified TF-IDF and word2vec. PLoS One. 2020; 15(3): e0230442. https://dx.doi.org/10.1371/journal.pone.0230442

26. Kayesh H, Islam MS, Wang J, Ohira R, Wang Z. SCAN: A shared causal attention network for adverse drug reactions detection in tweets. Neurocomputing. 2022; 479: 60–74. https://dx.doi.org/10.1016/j.neucom.2022.01.019

27. Rabani ST, Khan QR, Khanday A. Detection of suicidal ideation on Twitter using machine learning & ensemble approaches. Baghdad Sci J. 2020; 17(4): 1328. https://dx.doi.org/10.21123/bsj.2020.17.4.1328

28. Filieri R, Galati F, Raguseo E. The impact of service attributes and category on eWOM helpfulness: An investigation of extremely negative and positive ratings using latent semantic analytics and regression analysis. Comput Human Behav. 2021; 114: 106527. https://dx.doi.org/10.1016/j.chb.2020.106527

29. Hassan FH, Omar MA. Recurrent Stroke Prediction using Machine Learning Algorithms with Clinical Public Datasets: An Empirical Performance Evaluation. Baghdad Sci J. 2021; 18(4 Supplement). https://dx.doi.org/10.21123/bsj.2021.18.4.(Suppl.).1406

30. AL-Ani MM, Omar N, Nafea AA. A Hybrid Method of Long Short-Term Memory and Auto-Encoder Architectures for Sarcasm Detection. J Comput Sci. 2021; 17(11): 1093–8. https://dx.doi.org/10.3844/jcssp.2021.1093.1098

31. Lee K, Qadir A, Hasan SA, Datla V, Prakash A, Liu J, et al. Adverse drug event detection in tweets with semi-supervised convolutional neural networks. Proc 26th int conf world wide web. ACM Press. 2017: 705–14. https://dx.doi.org/10.1145/3038912.3052671 .

32. Cocos A, Fiks AG, Masino AJ. Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts. J Am Med Informatics Assoc. 2017; 24(4): 813–21. https://dx.doi.org/10.1093/jamia/ocw180

33. Wang CS, Lin PJ, Cheng CL, Tai SH, Yang YHK, Chiang JH. Detecting potential adverse drug reactions using a deep neural network model. J Med Internet Res. 2019; 21(2): e11016. https://dx.doi.org/10.2196/11016

الشبكة العصبية الاصطناعية والتحليل الدلالي الكامن لاكتشاف التفاعلات الضارة للأدوية

احمد عادل نافع   نازليا عمر   زهاء مبارك القفيل

مركز تكنولوجيا الذكاء الاصطناعي (CAIT) ، كلية علوم وتكنولوجيا المعلومات ، جامعة ماليزيا الوطنية (UKM) ، بانجي ، سيلانجور ، ماليزيا.

الخلاصة:

التفاعلات الدوائية الضارة (ADR) هي معلومات مهمة للتحقق من وجهة نظر المريض بشأن دواء معين. تم أخذ تعليقات ومراجعات المستخدمون المنتظمة في الاعتبار أثناء عملية جمع البيانات لاستخراج تأثيرات ال ADR عندما أبلغ المستخدم عن تأثير جانبي بعد تناول دواء معين. في الأدبيات، ركز معظم الباحثين على تقنيات التعلم الآلي لاكتشاف ADR. تعمل هذه الطرق على تدريب نموذج التصنيف باستخدام بيانات المراجعات الطبية. ومع ذلك، لا يزال هناك العديد من المشكلات الصعبة التي تواجه استخراج ال ADR، وخاصة دقة الكشف. الهدف الرئيسي من هذه الدراسة هو اقتراح الشبكات العصبية الاصطناعية (ANN) مع التحليل الدلالي الكامن (LSA) للكشف عن ال ADR. تظهر النتائج فعالية استخدام LSA مع ANN في استخراج ADR.

الكلمات المفتاحية: التفاعلات الدوائية الضارة، الشبكة العصبية الاصطناعية، التصنيف، التعلم العميق ، التحليل الدلالي الكامن.