ORIGINAL ARTICLE

A Highly Accurate Adverse Drug Reactions (ADR) Detection from Medical Forum Comments Using Long Short-Term Memory Networks

Anjali Basagodu Veeresh [1*], Ravikumar Guralamata Krishnegowda [2], Shashikala Salekoppalu Venkataramu [2]

[1] *Research Scholar Computer Science & Engineering Department, BGS Institute of Technology ACU, BG Nagar, Karnataka, India*

[2] *Research Scholar Computer Science & Engineering Department, BGS Institute of Technology ACU, BG Nagar, Karnataka, India*

*Corresponding Author: Anjali Basagodu Veeresh
Email: anjalibv.sampath@gmail.com

Received: 21 March 2023 / Accepted: 01 August 2023

Abstract

Purpose: Adverse Drug Reactions (ADR) classification is useful in modern medical diagnostics and related applications. ADR is an example of how medical information is frequently accessible on social media platforms for healthcare, where people can share their experiences with treatments on desktop computers and mobile devices. Many researchers are interested in gathering valuable medical data from social media for the ADR system training and classification process.

Materials and Methods: This research explores the effects of three aspects on recognizing ADR mentions in social media for the medical field and proposes a deep neural network of Long Short-Term Memory (LSTM) neural networks to do so. The comments are collected from various social media platforms to implement the ADR system with proper training and testing processes. The texts from the dataset are initially preprocessed by using a data filtering and clustering process to remove the input data's redundant information to increase the training process's quality. Characteristic features, such as semantic features and text statistics, are extracted from the input text using the American Standard Code for Information Interchange (ASCII) array. Further, the features are converted and fed to LSTM networks for training and validation.

Results and Conclusion: This work is evaluated using two datasets, CODEC, and ADR Corpus datasets are used to evaluate the performance of the proposed ADR technique via multiple angles. Via extensive experiments, this work achieved 99.79 accuracy, 98.37 sensitivity, 97.63 specificity, 99.72 precision, 98.39 recall, 97.62 F1-score for the CODEC dataset, 98.16 for accuracy, 99.19 for sensitivity, 98.49 for specificity, 99.49 for precision, 96.72 for recall, and 93.16 for F1-score for ADR corpus, respectively.

Keywords: Adverse Drug Reactions; Medical Information; Long Short-Term Memory; American Standard Code for Information Interchange; Sensitivity Evaluation.

DOI: https://doi.org/10.18502/fbt.v11i4.16508

1. Introduction

A negative, unanticipated outcome brought on by taking medication is known as an ADR. ADRs can happen after taking a single dose, over a lengthy period, or combining two or more medications. The phrase "side effect" has a different connotation because it can have positive and negative consequences. Any unanticipated or unwarranted occurrence while a drug is being used is an Adverse Event (AE), regardless of whether the event is connected to the drug's delivery. The ADRs are only one example of the kind of helpful medical information that is frequently accessible on social media platforms for healthcare, where people can share their own experiences with treatments on desktop computers and mobile devices. There are two types of adverse drug reactions: those that affect all patients but occur at varying dosages in various patients and those that only affect some patients but not all, regardless of the dose. There is a misconception that immunological responses, such as anaphylaxis, are dose-independent; nevertheless, the dose dependence may not be readily apparent within the therapeutic dosage range [1].

The response in affected patients will inevitably rely on the dose. Social media platforms (like Twitter) emerged during the past ten years and changed online networking and communication. These platforms are used for real-time trend tracking, information retrieval, and disease surveillance. One widely used social media platform is Twitter, which may be useful for real-time ADE detection. However, finding ADEs in Twitter tweets is not without its difficulties. For instance, (1) the absence of ADE tweets in the real-world Twitter stream, (2) the use of common language to express medical diseases, and (3) the occurrence of side effects such as conditions and medications in the same tweet without necessarily indicating an ADE [2].

Examining social media comments is one technique for early event detection. Examples include predicting whether users will remain on or quit health discussion boards (like DailyStrength and HealthBoards) and examining their motivations. As continuing participation in these forums may benefit patients and doctors, this has shown to be a promising area. Other instances include using Facebook to reveal drug usage, Twitter to track misconduct, and Facebook to use smoking cessation practices. Additionally, social media can give researchers access to particular types of data, such as a person's age, country, gender, and geolocation, that are typically unavailable due to data protection laws [3].

Additional features of social media language further constrain lexicon matching's applicability as an ADR detection technique. For instance, the language used on social media is informal, using slang terms and expressions (such as "feeling like crap") and containing frequent misspellings and errors in grammar (such as "dis Adderall has me sweating"). Additionally, symbols and abbreviations communicate semantic information, such as "lol" and emoticons [4]. ADR detection is used as a supervised machine-learning sequence-labeling problem to get around these issues, enabling the learning approaches to consider the input word's context. This is generally done through natural language processing, which tags each token (i.e., contiguous letter sequence, which is usually comparable to a word) with a named entity tag (e.g., person). For ADR detection, tokens might be recognized as a part of an unfortunate occurrence. The most effective ADR sequence labeling uses Conditional Random Field (CRF) models [5]. CRFs are constrained by the input they receive because the model only considers the target. Deep learning techniques are commonly used in various medical-related applications to perform prediction and classifications [6]. Various biomedical imaging-based applications are successfully designed and validated using complex image datasets for medical diagnostic-related applications [7 - 8].

This work performs highly accurate ADR detection using an LSTM classifier with multiple extracted features. To improve the quality of the training process, the texts from the dataset are first preprocessed using a data filtering and clustering technique. In the preprocessing stage, unwanted text characters and keywords are removed. Various features, including semantics, text statics, and ASCII array were extracted and fed to the LSTM for training and validation. Social media comments are extracted from various networks to study the performance of the proposed ADR implementation process. The sensitivity evaluation technique is applied to the classified labels to evaluate the performance of the proposed techniques.

The remainder of this paper is structured as follows. Section II summarizes some of the major work previously implemented to perform the ADR process. Section III details the proposed method using all mathematical terms for preprocessing, feature extraction, and classification. The performance of the suggested approach with the mathematical description of quality metrics is explained in Section IV. Section V concludes the paper with detailed outcomes obtained from the research.

1.1. Literature Review

Previously, several techniques were proposed to perform highly accurate ADR implementation. Various techniques used different preprocessing and feature extraction algorithms to achieve high accuracy. The major challenges in ADR classification are computational time and accuracy. Some techniques achieved acceptable accuracy with high computational complexity. Other techniques achieved moderate accuracy with less computation time. The computational time and accuracy should be balanced to obtain high performance of the ADR classification. The following session describes detailed information about the previous works performed.

Elena and Sergey proposed extracting adverse drug reactions from user reviews: together with a bidirectional LSTM-based recurrent neural network, CRF uses the scores that were recovered by this neural network. We compared this method to cutting-edge neural models on a sample ADR extraction dataset and discovered that the outcomes were much better. Additionally, adding a character-level model to input embeddings made additional gains. As a result, the final model effectively incorporates three different NLP statistical modeling methodologies [9].

Mert *et al.* proposed a technique for finding mentions of ADR entities in drug labels and normalizing them using the Medical Lexicon for Regulatory Activities (MedDRA) vocabulary utilizing machine learning and algorithms. The machine learning technique is based on a recently established deep learning architecture. It combines CRF, Convolutional Neural Networks (CNN), and Bidirectional Long Short-Term Memory (Bi-LSTM). Based on an improvement to the text-mining engine, SciMiner, the rule-based method converts the discovered ADR mentions to MedDRA phrases [10].

Kathy *et al.* proposed that combining many semi-supervised CNN models was recommended for categorizing ADE in tweets, especially employing a variety of unlabeled input categories to create the models. When only a portion of the available unlabeled data is employed, semi-supervised CNN models beat supervised classification models by a +9.9% F1-score evaluated the models using the Twitter data set from the PSB 2016 Social Media Shared Task [11].

Chuhan *et al.* proposed a neural network technique to simultaneously find tweets involving drug names or negative drug effects. An amalgamation A hierarchical tweet representation approach is used to learn language models from characters and then build depictions of tweets from words to lessen the impact of frequent misspellings and user-created abbreviations in tweets. Consider employing a multi-head self-attention technique to depict word exchanges in tweets further to better portray tweet contexts. To offer more informative tweet representations, incorporate the additive attention strategy while choosing informative terms [12].

Liliya and Mikhail proposed a CNN-based binary classification approach to the problem of ADR detection in Twitter data. For better word embeddings, various preprocessing methods were used. Finally, a CNN is given these embeddings to train the ADR classifier. The Google News word embeddings produced the best results and achieved an accuracy score of 90.4% on the test data and an ADR F-score of 54.23%, demonstrating the applicability of deep learning methods to these kinds of applications [13].

The major challenge in ADR classification is computational time and flexibility. The acceptable value of computational time is one of the important factors in the ADR classification. Previously, CNN, LSTM, and Bi-LSTM techniques were proposed and implemented with direct and feature extraction techniques. The direct feeding of inputs in the LSTM increases the computational complexity due to a huge amount of data processing in the LSTM. Hierarchical tree implementation was also proposed to perform the ADR classification by maintaining moderate accuracy. Classical ADR techniques were designed in such a way as to maintain moderate computational accuracy with minimum time. The proposed method uses multiple-feature extraction to train the LSTM to obtain maximum accuracy and minimize the computational time by minimum input size by performing a significant feature extraction process.

The major objectives of this work are stated below:

1. An improvement in the classification accuracy of ADR by performing accurate preprocessing and extracting significant and powerful features from the input text array.

2. To reduce the complexity of the detection process, which improves the potential of the process of real-time testing.

3. To improve the flexibility of the design for various datasets to produce highly accurate results after training with commonly available existing datasets.

2. Materials and Methods

2.1. Proposed Accurate adverse Drug Reactions (ADR) Detection Using Multiple Feature Based LSTM

2.1.1. Block Diagram

The texts that are part of the dataset in this work are split into sentences, stripped of keywords, and devoid of special characters so that characteristics can be retrieved from them. Some features include semantic characteristics, text statics, and ASCII arrays. Carry out the matrix array-to-text array conversion. After that, an LSTM network was used to categorize the data. The LSTM classified performance is evaluated. Figure 1 shows the block diagram of the proposed method.

2.1.2. Preprocessing

In this work, a cutting-edge CNN architecture that divides a paragraph into smaller pieces, such as phrases

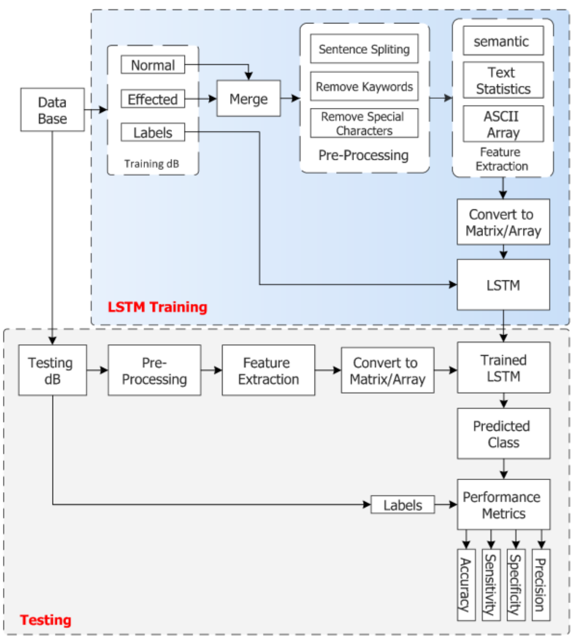

Figure 1. Block diagram of the Proposed ADR classification with LSTM

or words, is known as sentence splitting. Then, each unit is regarded as a separate sentence [14]. The main idea behind sentence splitting is to examine the tokens or smaller units that make up each paragraph to comprehend the meaning of the text as a whole. It is also known as Tokenization. Eliminate special characters such as @! /, *, $, etc. Remove words like, at, of, the, etc. [15-16].

2.1.3. Feature Extraction

A: Semantic Feature

When attempting to extract a word's semantic qualities from a text, it is important to consider the context in which it is used. Word2Vector, which trains the right word vector based on the context of the word in the text, is essential for extracting the semantics of words [17]. Word representation is transformed into a space vector using a technique called Word2Vector. To train a corpus, it largely uses the idea of deep learning by mapping each word's context to a distinct N-dimensional vector [18]. The semantic features enable the computer to communicate and recognize each word's semantic properties.

B: Text Static Features

Standard Deviation: A measure of the variance in the distribution of data collection in statistics is the standard deviation. Higher standard deviations are spread over a wider range, whereas smaller standard deviations are typically set as means. The formula for the sample standard deviation is (Equation 1):

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2} \tag{1}$$

Mean: The mean of a set of observed data is determined by summing the numerical values of all observations and dividing the result by the overall number of observations (Equation 2):

$$\bar{x} = \frac{1}{n} \left(\sum_{i=1}^{n} x_i \right) = \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{2}$$

Kurtosis: The mean of a set of observed data is determined by summing the numerical values of all observations and dividing the result by the overall number of observations (Equation 3).

$$Kurtosis[X] = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{E[(X-\mu)^4]}{(E[(X-\mu)^2])^2} = \frac{\mu_4}{\sigma^4} \tag{3}$$

Skewness: The asymmetry of a real-valued random variable's probability distribution concerning its mean is described by the statistic known as skewness in probability theory and statistics (Equation 4).

$$\tilde{\mu}_3 = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3} = \frac{E[(X-\mu)^3]}{(E[(X-\mu)^2])^{3/2}} = \frac{k_3}{k_2^{3/2}} \quad (4)$$

Moment: In both statistics and mechanics, the moment idea is applied. If the function represents mass, then the total mass is the zeroth moment. The n-th moment of a real-valued continuous function f(x) of a real variable about a value c is (Equation 5):

$$\mu_n = \int_{-\infty}^{\infty} (x-c)^n f(x)\ dx \quad (5)$$

Energy:

$$EG = \sum_i^{N_G}\sum_j^{N_G}\{p(i,j)\}^2 \quad (6)$$

***Entropy (EN)*:**

$$EN = -\sum_i^{N_G}\sum_j^{N_G} p(i,j)\, log(p(i,j)) \quad (7)$$

$p(i,j)$ in a normalized matrix N_G is after that. Quantized image number of distinct grey levels.

Inertia (IN):

$$IN = -\sum_i^{N_G}\sum_j^{N_G}(i-j)^2\, p(i,j) \quad (8)$$

Correlation (CO):

$$CO = \frac{\sum_i^{N_G}\sum_j^{N_G}(i-\mu_x)(j-\mu_y)\ p(i,j)}{\sigma_x\sigma_y} \quad (9)$$

Where $\mu_x, \mu_y, \sigma_x, and\ \sigma_y$ are the means and standard deviations of $p_x p_y$.

Inverse Difference Moment (IDM):

IDM is written as:

$$idm = \sum_i^{N_G}\sum_j^{N_G}\frac{1}{1+(i-j)^2}\ p(i,j) \quad (10)$$

Difference Entropy (DE):

$$DE = -\sum_{K=0}^{N_G-1} P_{x-y}(k) log_2\, P_{x-y}(k) \quad (11)$$

Homogeneity:

$$homo = \sum_{i,j=0}^{G-1}\frac{p(i,j)}{1+(i+`j)^2} \quad (12)$$

Angular Second Moment:

$$asm = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\{p_\theta(i,j)\}^2 \quad (13)$$

Variance:

$$var = \sum\nolimits_{i,j} p_{i,j}(i-\mu_i)^2 \quad (14)$$

Difference Variance:

$$dffv = -\sum_{i=0}^{N_g-1}(i-f_6)^2 p_{x-y}(i) \quad (15)$$

Where $f_6 = \sum_{\mathrm{i,j}} \mathrm{p_{i,j}}|\mathrm{i}-j|$

Fractal Dimension:

$$fd = \frac{log(c) - log\ (N(r))}{log\ (r)} \quad (16)$$

C: ASCII Array

The 256 ASCII characters were used as a feature vector to record the relative frequencies of the individual characters in the payload to extract the features. The ASCII is rather cryptic because it was initially created for use with teletypes and is routinely used contrary to its intended usage. ASCII is a character set that uses a 7-bit system and has 128 characters. It includes the capital and lowercase letters A through Z, the numerals 0 through 9, and special characters.

D: LSTM

LSTM artificial neural networks are used in both deep learning and artificial intelligence. Unlike traditional feedforward neural networks, LSTM has feedback connections. The LSTM neural network was among the most widely used in the 20th century. In the terminology of the LSTM, a typical RNN is called possessing both "long-term memory" and "short-term memory." A cell, an input gate, an output gate, and a forget gate comprise an LSTM unit [19-20]. The three gates regulate the flow of information into and out of the cell, which stores values throughout time.

Given the possibility of latencies of variable durations between significant occurrences in a time series, LSTM networks are especially well suited for categorizing, analyzing, and generating recommendations for time series analysis. LSTMs were created to address the issue of disappearing gradients during routine RNN training. Because of their relative lack of compassion for gap length, LSTM outperforms RNNs, hidden Markov models, and other sequence evolutionary computations in many circumstances. Figure 2 shows the architecture of LSTM.

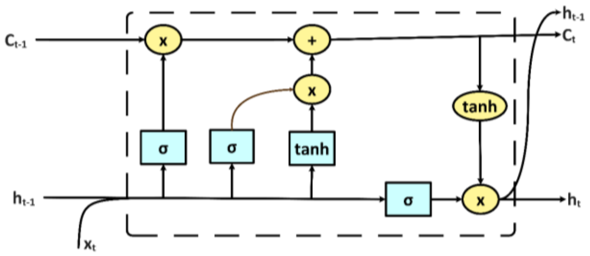

Figure 2. Architecture of LSTM

Algorithm 1. Algorithm of Proposed method

Input: Phrases(P_k)
Output: Effect types

For k=0 to N
  Read Phrases P_k,
  P_P = Eliminate (P_k(@, !, #,∗...))
  P_s = Static (P_p)
  P_A = ASCII array (P_A)
  LSTM = TrainLSTM(P_A)
  Store
EndFor
For k=0 to N
  Read Testing Phrases P_k
  PhrasesClasses = TestLSTM(LSTM , P_A)
EndFor

3. Results and Discussion

The proposed work is implemented using a Python programming language in an Anaconda Jupiter notebook. A 64-bit Windows 10 Pro PC was installed with an Intel i7-7700 processor and four cores clocked at 3.60 GHz, with 12 gigabytes of main memory. The datasets are stored in a text file in local system memory.

3.1.Dataset

The CSIRO Adverse Drug Event Corpus (CADEC) is a brand-new, comprehensive corpus of medical forum postings on adverse drug events reported by patients (ADEs). The corpus, which includes material that frequently veers from conventional English grammar and punctuation norms and is mostly written in colloquial language, is drawn from posts on social media [21]. Drugs, side effects, symptoms, and illnesses are mentioned in annotations and associated concepts from restricted vocabularies [22].

ADE-Corpus-V2 Dataset is an annotation guideline, multi-stage annotations, evaluating inter-annotator agreement, and a clinical terminologist's final review of the annotations to assure the quality of the annotations [23]. The uncover possible pharmaceutical side effects from patient stories on social media, this corpus may be utilized for research on information extraction or, more generally, text mining [24]. Table 1 shows the description of the CADEC dataset. In this work, 1321 posts are used. The 1250 posts contain text and 101486 words.

Table 1. Description of the CODEC dataset

	Corpus
No. of Posts	1321
No. posts with text	1250
No. words	101,486

Table 2 shows the description of the dataset. The CADEC entities are drug, ADR, disease, symptoms, and findings. ID, drug, the effect is the ADE corpus v2 entities.

3.2. Performance Metrics

This section evaluates the performance of the proposed method. Even for human annotators, it might be challenging to establish the borders of expressions. Thus, follow these procedures and carry out the experimental evaluation.

The proportion of offensive tweets correctly classified as offensive (TP) and non-tweets properly classified as non-offensive (TN) over the entire testing set are examples of accurate measurements (Equation 17).

$$accuracy = \frac{TP + TN}{T} \quad (17)$$

Where T is the total population = $TP + TN + FP + FN$.

Table 2. Entity Description of dataset

	Entity	Example	Annotated word
CODEC	Drug	I must be addicted to Diclofenac	Diclofenac
	ADR	Sometimes causes drowsiness	drowsiness
	Disease	after three years of using Ativan to control anxiety and anger.	anxiety & aggression
	Symptom	My heart was racing, etc.	Heart racing
	Finding	Any negative side effect, illness, symptom, or another clinical idea that may be classified in any of these categories but was not directly experienced by the reporting patient is referred to as a clinical finding.	Which one it belongs to is unclear, according to the annotator.
ADE corpus v2	Id	10030778	-
	Text	Intravenous azithromycin-induced ototoxicity	ototoxicity
	Drug	azithromycin	azithromycin
	Effect	ototoxicity	ototoxicity

A test's sensitivity (Se) describes its ability to identify people with true positives. A common name for it is the True Positive Rate (TPR). Mathematically, it can be expressed as (Equation 18):

$$Sensitivity = \frac{TP}{TP + FN} \tag{18}$$

Where TP = True Positives, FN = Number of False Negatives).

Specificity is the capacity of a test to correctly distinguish tweets that do not have the side effects (Sp). It is also known as the True Negative Rate (TNR) (Equation 19).

$$Specificity = \frac{TN}{TN + FP} \tag{19}$$

Out of all occurrences in the testing set that were either properly or mistakenly identified as offensive, precision is used to determine the proportion of correctly classified offensive tweets [25] (Equation 20).

$$precision = \frac{TP}{TP + FP} \tag{20}$$

Recall measures the proportion of offensive tweets in the testing set that were classified as the same [26] (Equation 21).

$$recall = \frac{TP}{TP + FN} \tag{21}$$

The harmonic mean of recall and precision together make up the F-score [27], which is calculated as (Equation 22):

$$f - score = 2 * \frac{precision * recall}{precision + recall} \tag{22}$$

The corpus was divided into two unique datasets, 25% of which (375 reviews, 2356 phrases, and 1837 ADRs) were utilized for testing and 75% of which (training data) (a total of 875 reviews, 5264 sentences, and 3933 ADRs).

Table 3 shows the performance of the proposed method. The CADEC Dataset achieves an accuracy of 99.79, 98.37 of sensitivity, 97.63 of specificity, 98.39 of precision, 98.39 of recall, and 97.62 of F1 score.

Table 3. Performance of the proposed method

Dataset	Accuracy	Sensitivity	Specificity	Precision	Recall	F1
CODEC	99.79	98.37	97.63	99.72	98.39	97.62
ADE corpus	98.16	99.19	98.49	99.49	96.72	93.16

ADE corpus achieves 98.16 of accuracy, 99.19 of sensitivity, 98.49 of specificity, 99.49 of precision, 96.72 of recall, 93.16 of F1 score.

Table 4 shows the performance of the proposed method training and testing. In the CODEC dataset, 75 percent training images and 25 testing images of 99.79 accuracy, 98.37 sensitivity, 97.63 specificity, 98.39 precision, and 98.39 recall. In the ADE corpus dataset, 75 percent of training images and 25 testing images achieves 98.16 accuracy, 99.19 sensitivity, 98.49 specificity, 99.49 precision, 96.72 recall, 93.16 F1 score.

Table 5 shows the performance of the proposed method compared with the existing method. Method [1] has 87.81 precision, 88.81 Recall, and 88.30 F1-score. Method [14] returns 74.47 precision, 64.96 Recall, and 69.39 F1-score. [15] achieves 88.8 precision, 85.5 Recall, and 87.26 F1-score. This work achieves 99.79 accuracy, 99.72 precision, 98.39 Recall, and 97.62 F1-score.

Figure 3 shows the performance metrics with different techniques. The red blue represented as method [1] orange bar is noted as [14]. Ash color is represented as [15]. Yellow color is represented in this work. This work achieves the highest performance.

Figure 4 shows the performance metrics for two different datasets. The CADEC dataset has the highest accuracy, precision, recall, and F1 score. ADE corpus achieves the highest sensitivity and specificity compared to the CADEC dataset. Figure 5 shows the confusion matrix of the ADR corpus. Figure 6 shows the confusion matrix of the CADAC. The healthy comments are 95 true positives, mild side effect comments are 98 percent true positives, and 96 percent true positive cases.

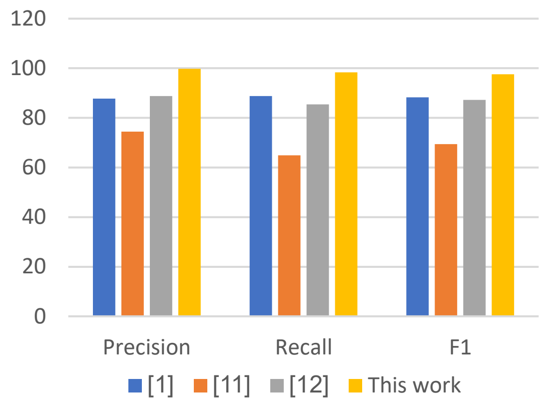

Figure 3. Performance metrics with different techniques

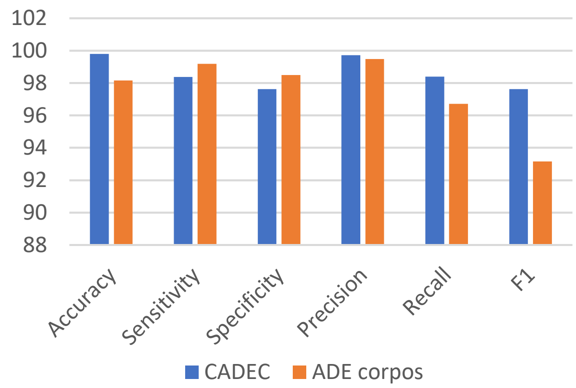

Figure 4. Performance metrics for two different datasets

Table 5. Performance of the proposed method compared with the existing method

Methods	Accuracy	Precision	Recall	F1
[1]	-	87.81	88.81	88.30
[11]	-	74.47	64.96	69.39
[12]	-	88.8	85.5	87.26
This work	99.79	99.72	98.39	97.62

Table 4. Performance of the proposed method based on training and testing

	Training	Testing	Accuracy	Sensitivity	Specificity	Precision	Recall
CODEC	35	65	42.36	44.49	58.63	43.19	58.63
	30	70	67.88	47.97	46.31	68.97	69.46
	50	50	82.29	83.68	88.56	56.69	86.35
	75	25	99.79	98.37	97.63	99.72	98.39
ADE corpus	35	65	53.65	57.36	58.97	60.37	63.97
	30	70	62.17	66.98	64.89	72.19	69.34
	50	50	72.36	72.97	82.49	78.94	79.31
	75	25	98.16	99.19	98.49	99.49	96.72

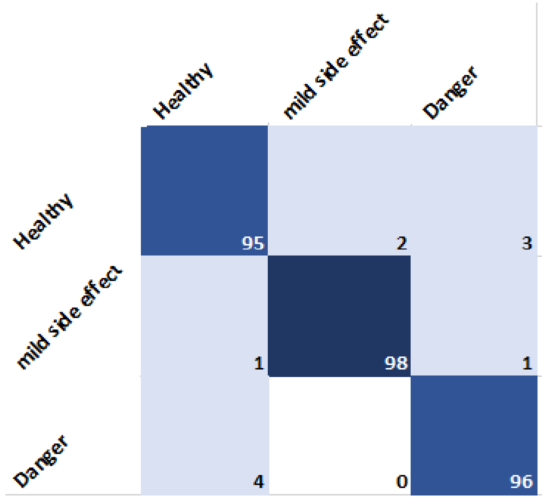

Figure 5. Confusion matrix of the ADR corpus

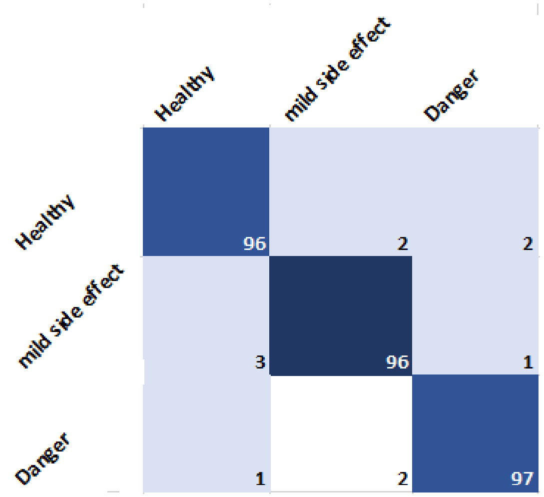

Figure 6. Confusion matrix of the CADAC

4. Conclusion

This work concentrated on the issue of automatically classifying phrases in the text to find ADR mentions. In this work, the texts included in the dataset are preprocessed using the splitting sentences, removing keywords, and removing special characters so that features may be extracted from them. These features include semantics features, text statics, and ASCII arrays. Perform the conversion from the text array to the matrix array. The data was then categorized using an LSTM network. This work achieved 99.79 of accuracy, 98.37 sensitivity, 97.63 specificity 99.72 precision 98.39 recall, and 97.62 F1-score for the CADEC dataset. 98.16 accuracy, 99.19 sensitivity, 98.49 specificity, 99.49 for precision, 96.72 of recall, 93.16 F1-score for ADE corpus.

References

1- Hasham Ul Haq, Veysel Kocaman, and David Talby, "Mining adverse drug reactions from unstructured mediums at scale." in *Multimodal AI in healthcare: A paradigm shift in health intelligence*: *Springer*, (2022), pp. 361-75.

2- Anne Cocos, Alexander G Fiks, and Aaron J Masino, "Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts." *Journal of the American Medical Informatics Association*, Vol. 24 (No. 4), pp. 813-21, (2017).

3- IS Alimova and EV Tutubalina, "Entity-level classification of adverse drug reaction: A comparative analysis of neural network models." *Programming and Computer Software*, Vol. 45pp. 439-47, (2019).

4- Claudia Giardina *et al.*, "Adverse drug reactions in hospitalized patients: results of the FORWARD (Facilitation of Reporting in Hospital Ward) study." *Frontiers in pharmacology*, Vol. 9p. 350, (2018).

5- Jonathan A Galli, Aloknath Pandya, Michelle Vega-Olivo, Chandra Dass, Huaqing Zhao, and Gerard J Criner, "Pirfenidone and nintedanib for pulmonary fibrosis in clinical practice: tolerability and adverse drug reactions." *Respirology*, Vol. 22 (No. 6), pp. 1171-78, (2017).

6- Mehdi Gheisari *et al.*, "Deep learning: Applications, architectures, models, tools, and frameworks: A comprehensive survey." *CAAI Transactions on Intelligence Technology*, (2023).

7- Mustafa Ghaderzadeh, Mehrad Aria, Azamossadat Hosseini, Farkhondeh Asadi, Davood Bashash, and Hassan Abolghasemi, "A fast and efficient CNN model for B-ALL diagnosis and its subtypes classification using peripheral blood smear images." *International Journal of Intelligent Systems*, Vol. 37 (No. 8), pp. 5113-33, (2022).

8- Mustafa Ghaderzadeh, Farkhondeh Asadi, Ramezan Jafari, Davood Bashash, Hassan Abolghasemi, and Mehrad Aria, "Deep convolutional neural network–based computer-aided detection system for COVID-19 using multiple lung scans: design and implementation study." *Journal of Medical Internet Research*, Vol. 23 (No. 4), p. e27468, (2021).

9- Elena Tutubalina and Sergey Nikolenko, "Combination of deep recurrent neural networks and conditional random fields for extracting adverse drug reactions from user reviews." *Journal of healthcare engineering*, Vol. 2017(2017).

10- Mert Tiftikci, Arzucan Özgür, Yongqun He, and Junguk Hur, "Machine learning-based identification and rule-based normalization of adverse drug reactions in drug labels." *BMC bioinformatics*, Vol. 20 (No. 21), pp. 1-9, (2019).

11- Kathy Lee *et al.*, "Adverse drug event detection in tweets with semi-supervised convolutional neural networks." in *Proceedings of the 26th international conference on world wide web*, (2017), pp. 705-14.

12- Chuhan Wu, Fangzhao Wu, Zhigang Yuan, Junxin Liu, Yongfeng Huang, and Xing Xie, "MSA: Jointly detecting drug name and adverse drug reaction mentioning tweets with multi-head self-attention." in *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, (2019), pp. 33-41.

13- Liliya Akhtyamova, Mikhail Alexandrov, and John Cardiff, "Adverse drug extraction in twitter data using convolutional neural network." in *2017 28th International Workshop on Database and Expert Systems Applications (DEXA)*, (2017): *IEEE*, pp. 88-92.

14- Sunil Kumar Sahu and Ashish Anand, "Drug-drug interaction extraction from biomedical texts using long short-term memory network." *Journal of biomedical informatics,* Vol. 86pp. 15-24, (2018).

15- Bharath Dandala, Venkata Joopudi, and Murthy Devarakonda, "Adverse drug events detection in clinical notes by jointly modeling entities and relations using neural networks." *Drug safety,* Vol. 42pp. 135-46, (2019).

16- Jing Liu, Songzheng Zhao, and Xiaodi Zhang, "An ensemble method for extracting adverse drug events from social media." *Artificial intelligence in medicine,* Vol. 70pp. 62-76, (2016).

17- Hong-Jie Dai, Musa Touray, Jitendra Jonnagaddala, and Shabbir Syed-Abdul, "Feature engineering for recognizing adverse drug reactions from twitter posts." *Information,* Vol. 7 (No. 2), p. 27, (2016).

18- Zahra Rezaei, Hossein Ebrahimpour-Komleh, Behnaz Eslami, Ramyar Chavoshinejad, and Mehdi Totonchi, "Adverse drug reaction detection in social media by deep learning methods." *Cell Journal (Yakhteh),* Vol. 22 (No. 3), p. 319, (2020).

19- Sarthak Jain, Xun Peng, and Byron C Wallace, "Detecting Twitter posts with Adverse Drug Reactions using Convolutional Neural Networks." in *SMM4H@ AMIA*, (2017), pp. 72-75.

20- KM Deepa Lakshmi and A Prakash, "Evaluating the Adverse Drug Reactions Learning Framework Using Novel Text Mining." *International Journal of Applied Engineering Research,* Vol. 13 (No. 21), pp. 14983-96, (2018).

21- Sarvnaz Karimi, Chen Wang, Alejandro Metke-Jimenez, Raj Gaire, and Cecile Paris, "Text and data mining techniques in adverse drug reaction detection." *ACM Computing Surveys (CSUR),* Vol. 47 (No. 4), pp. 1-39, (2015).

22- Shashank Gupta, Manish Gupta, Vasudeva Varma, Sachin Pawar, Nitin Ramrakhiyani, and Girish Keshav Palshikar, "Multi-task learning for extraction of adverse drug reaction mentions from tweets." in *European Conference on Information Retrieval*, (2018): *Springer*, pp. 59-71.

23- Peng Ding, Xiaobing Zhou, Xuejie Zhang, Jin Wang, and Zhenfeng Lei, "An attentive neural sequence labeling model for adverse drug reactions mentions extraction." *Ieee Access,* Vol. 6pp. 73305-15, (2018).

24- Daniel S Budnitz, Nadine Shehab, Scott R Kegler, and Chesley L Richards, "Medication use leading to emergency department visits for adverse drug events in older adults." *Annals of internal medicine,* Vol. 147 (No. 11), pp. 755-65, (2007).

25- Robert P Schumaker, Michael A Veronin, Trevor Rohm, Matthew Boyett, and Rohit R Dixit, "A data driven approach to profile potential SARS-CoV-2 drug interactions using TylerADE." *Journal of International Technology and Information Management,* Vol. 30 (No. 3), pp. 108-42, (2021).

26- Eva-Lisa Meldau, "Deep neural networks for inverse de-identification of medical case narratives in reports of suspected adverse drug reactions." ed, (2018).

27- Nitin Ramrakhiyani and Girish Keshav Palshikar, "Multi-task Learning for Extraction of Adverse Drug Reaction Mentions from Tweets." in *Advances in Information Retrieval: 40th European Conference on IR Research, ECIR 2018, Grenoble, France, March 26-29, 2018, Proceedings*, (2018), Vol. 10772: *Springer*, p. 59.