



Article

Drug-Drug Interaction Extraction from Biomedical Text Using Relation BioBERT with BLSTM

Maryam KafiKang * and Abdeltawab Hendawi 

Department of Computer Science and Statistics, University of Rhode Island, Kingston, RI 02881, USA;
hendawi@uri.edu

* Correspondence: maryam_kafikang@uri.edu

Abstract: In the context of pharmaceuticals, drug-drug interactions (DDIs) occur when two or more drugs interact, potentially altering the intended effects of the drugs and resulting in adverse patient health outcomes. Therefore, it is essential to identify and comprehend these interactions. In recent years, an increasing number of novel compounds have been discovered, resulting in the discovery of numerous new DDIs. There is a need for effective methods to extract and analyze DDIs, as the majority of this information is still predominantly located in biomedical articles and sources. Despite the development of various techniques, accurately predicting DDIs remains a significant challenge. This paper proposes a novel solution to this problem by leveraging the power of Relation BioBERT (R-BioBERT) to detect and classify DDIs and the Bidirectional Long Short-Term Memory (BLSTM) to improve the accuracy of predictions. In addition to determining whether two drugs interact, the proposed method also identifies the specific types of interactions between them. Results show that the use of BLSTM leads to significantly higher F-scores compared to our baseline model, as demonstrated on three well-known DDI extraction datasets that includes SemEval 2013, TAC 2018, and TAC 2019.

Keywords: drug-drug interactions; relation extraction; natural language processing; bidirectional long short-term memory; relation biobert; deep learning



Citation: KafiKang, M.; Hendawi, A. Drug-Drug Interaction Extraction from Biomedical Text Using Relation BioBERT with BLSTM. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 669–683. <https://doi.org/10.3390/make5020036>

Academic Editors: S. Ejaz Ahmed, Abdulkadir Hussein and Abbas Khalili

Received: 17 May 2023
Revised: 3 June 2023
Accepted: 8 June 2023
Published: 10 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Drug-drug interactions (DDIs) refer to the effects produced when two or more drugs interact, potentially impacting the behavior of the drugs [1]. In certain circumstances, DDIs can cause adverse drug reactions (ADRs), which pose serious health hazards and life-threatening issues [2]. The use of multiple drugs simultaneously increases the risk of DDIs, which can endanger patients' lives and lead to fatalities. DDIs pose a significant bottleneck for drug administration and patient safety, making them a critical factor affecting drug-related side effects and patient health [3].

According to U.S. center reports, ADRs cause 300,000 deaths each year in the U.S. and Europe [4]. Furthermore, 10% of individuals take five or more drugs simultaneously, with 20% of the elderly population taking at least ten drugs at the same time [5], dramatically increasing the risk of ADRs. Due to the significance of DDIs in providing vital information to patients, medical researchers and public health physicians must possess accurate and up-to-date knowledge of DDIs in order to prescribe drugs safely and effectively.

With the increasing use of drugs, it has become essential to maintain databases that store comprehensive drug information. However, keeping these databases up to date with the exponential growth of biomedical literature is a significant challenge [6,7]. Despite the integration of databases such as DrugBank [8], Therapeutic Target DB [9], and PharmGKB [10], which provide drug information and DDIs to medical researchers and scientists, a substantial amount of DDIs information remains locked in biomedical articles rather than being incorporated into databases. As a result, there is a need to develop automatic methods for extracting DDIs information from these resources. Automatic DDIs extraction

has the potential to greatly benefit the pharmaceutical industry by reducing the time spent by healthcare professionals in reviewing medical literature. Therefore, the development of automatic methods for extracting DDIs from texts is necessary.

DDIs extraction is a relation classification task that involves categorizing pairs of drug entities into predefined categories within the context of the sentence. Unlike typical text classification tasks, DDIs extraction requires the model to have knowledge of the drug entities to accurately learn and perform the classification task.

Numerous techniques have been proposed to extract drug-drug interactions (DDIs), which can be classified into two categories: pattern-based methods and feature-based machine learning methods. Pattern-based methods rely on manual patterns for DDIs classification, which makes them time-consuming and laborious and need a set of domain specific knowledge [11]. On the other hand, feature-based machine learning methods have demonstrated superior performance compared to pattern-based approaches over the past decades. However, these methods heavily rely on specific features, limiting their ability to capture other significant patterns within texts.

In contrast, deep learning approaches enable models to automatically learn data representations [12]. To improve the accurate representation of semantic information in text-related tasks, some methods incorporate techniques from natural language processing (NLP) [13] and employ word vector models such as Glove [14] and Word2vec [15] to convert each processed token (i.e., the smallest unit of text processing) into a vector representation. The remarkable performance of BERT (bidirectional encoder representation of transformers) [16], as well as similar pre-training models, such as SciBERT [17] and BioBERT [16] in the field of NLP, has led to their successive application in DDI extraction. Particularly, BioBERT, a biomedical language representation model pre-trained on an extensive biomedical corpus, accurately captures the semantic information embedded within medical-related texts.

We illustrate the overview of the proposed method in Figure 1. In our model, we employed R-BioBERT with BLSTM to identify and classify DDIs in biomedical texts. Moreover, our model goes beyond traditional approaches by extracting DDIs through comprehensive analysis of the relationship between two drugs within a sentence, surpassing the limitations of solely relying on drug names or targets. Specifically, our proposed model incorporates the information from large-scale raw texts by using Relational BioBERT then uses BLSTM to classify DDIs in sentences.

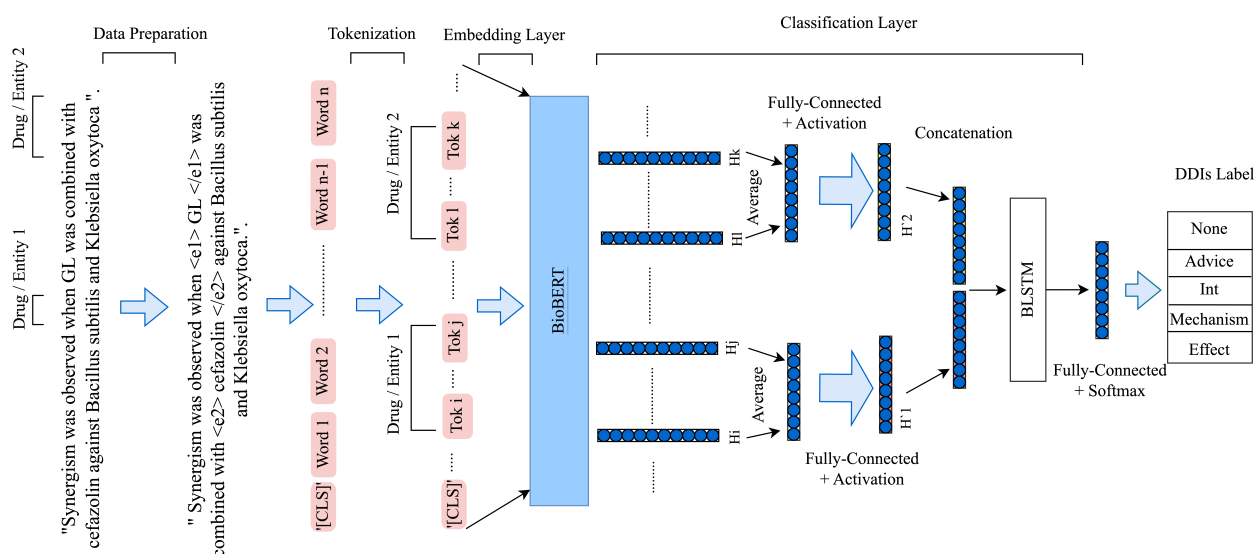


Figure 1. The model architecture.

We evaluated our method on the SemEval 2013, TAC 2018, and TAC 2019. Experimental results show that BLSTM boosts the performance of the baseline model (R-BioBERT).

Our model (R-BioBERT with BLSTM) achieved an F1-Macro of 83.32% on SemEval 2013, 80.23% on TAC 2019, and 60.53% on TAC 2018 DDIs Extraction. These findings indicate that our model outperforms the baseline model (R-BioBERT).

The main contribution of this work can be summarized as follows:

- Our study proposes a novel approach that leverages the power of integrating BLSTM and Relation BioBERT to accurately extract drug-drug interactions (DDIs) and classify their respective types of relationships.
- To evaluate the efficacy of our proposed model, we conducted experiments on three distinct datasets: SemEval 2013, TAC 2018, and TAC 2019 DDIs Extraction, all of which involve drug-drug interactions (DDIs) extraction tasks. Our experimental results demonstrate that our proposed method (is R-BioBERT with BLSTM) outperforms the baseline model.

The subsequent sections of this paper are organized as follows: In Section 2, we review related works in the field of drug-drug interactions (DDIs) extraction. Section 3 presents a detailed literature review pertaining to the subject matter of this study. Section 4 presents the dataset and our proposed method, which utilizes R-BioBERT with BLSTM. In Section 5, the experiment setup and our evaluations metrics are explained. Section 6 reports and analyzes the experimental results, demonstrating the superior performance of our method compared to the baseline. Finally, Section 7 provides conclusions and summarizes the contributions of this work.

2. Related Works

DDIs extraction entails discovering semantic relationships between pairs of drugs. Supervised methods [18–20] are the predominant approaches used in this task employing deep learning techniques. Recently, the application of recurrent neural networks (RNNs) [21], convolutional neural networks (CNNs) [22], and recursive neural networks (recursive NNs) in DDIs extraction has demonstrated their ability to acquire significant information and outperform conventional machine learning techniques.

Convolutional Neural Networks (CNNs) are a powerful deep learning technique that has gained significant attention in various real-world applications, such as image classification [23], object detection [24], and several engineering applications [25]. CNNs are also applicable to natural language processing tasks, such as sentiment analysis [26], search query [27], and semantic parsing [28]. CNNs have been applied in DDIs extraction tasks. The first application of CNNs in DDIs extraction was developed by Liu, Shengyu et al. [29]. Asada et al. [30] proposed a method that combined attention mechanisms with CNN, which outperformed the CNN-based model [29] in DDIs extraction. Some studies, such as [31,32], have increased the depth of the CNN architecture by creating deeper networks. Additionally, Asada, Masaki, et al. [9] proposed a model that applied CNNs with a graph that encoded sentences and molecular drug pairs in the DDIs extraction task. MCCNN [33] introduced a method that utilizes distributed word embedding and a multichannel convolutional neural network for biomedical relation extraction. Finally, Sun et al. [34] developed a method based on Bidirectional Long Short-Term Memory (BLSTM) to capture semantic knowledge from texts and a CNN to extract sentence features.

Recurrent Neural Networks (RNNs) are a popular deep learning method that excels at capturing sentence features, making them particularly well-suited for natural language processing tasks. Kavuluru et al. [35] developed a method that combines an original word-based RNN with a character-based RNN. D. Huang et al. [36] proposed a two-stage LSTM model that uses an SVM model to recognize negative and positive DDIs, and an LSTM model to classify DDIs into a specific category. Z. Yi and S. Li [37] proposed the 2ATT-RNN, which includes two attention layers that are a word-level attention layer and a sentence-level attention layer. Another RNN-based method is the joint AB-LSTM model, which is proposed by Sahu et al. [20] for DDIs extraction. Zhou et al. [38] presented the PM-BLSTM model, which incorporates a position-aware attention mechanism to encode the relative position information of the target entities with the hidden states of the BLSTM layer.

Contextualized embedding-based methods have gained significant attention recently, and have shown promising results in various NLP tasks [39]. Deep transformer-based methods are trained using contextualized embeddings on large text data. BERT, a pre-trained language model, has been extensively applied in many NLP tasks due to its bidirectional encoder transformer architecture, which captures richer context compared to other word embedding methods such as Glove and Word2vec. BERT has been shown to improve the performance of models by integrating contextual information in sentences. For example, Datta et al. [40] developed a BERT-based model for extracting DDIs from sentences. Similarly, Zaikis et al. [41] proposed a deep learning model based on the transformer architecture and the BERT language model for DDIs extraction tasks. Moreover, bio-specific BERT models such as BioBERT [42], SciBERT [10], and Med-BERT [43] have been applied in several DDIs extraction tasks such as [44,45] that combine BioBERT and SciBERT to obtain richer sequence semantic information. In addition, [46] proposed a method called IK-DDI that uses instance position embedding and key external text for DDI extraction tasks. Huang, Zhong, et al. [47] proposed a EMSI-BERT method to improve pre-trained BERT performance in DDI classification tasks using a drug entity dictionary, an Entity-Mask strategy, and a Symbol-Insert structure. The paper [48] proposes a neural network-based method using output-modified BioBERT and multiple entity-aware attentions for drug-drug interaction classification.

3. Literature Review

This section of the paper aims to establish a foundation for our proposed model by providing an overview of BERT, BioBERT, R-BioBERT, and BLSTM.

3.1. BERT Language Model

In NLP tasks, the use of language representation models has become essential in order to learn word representations from unlabeled texts. While previous language models, such as Glove and Word2vec, are context-free and focus on learning word representations without considering the context of words in sentences, recent language representation models, such as ELMO [49], Cove [50], OpenAI GPT [51], and BERT, are context-sensitive and emphasize on learning contextual word embeddings. These models have been proven effective in various NLP applications and are gaining popularity within the field.

BERT, which stands for Bidirectional Encoder Representations from Transformers, is a pre-trained language representation model that utilizes a context-sensitive approach for word representation [16]. BERT is a general-purpose language model that has been trained on massive datasets, including English Wikipedia and Books Corpus, to obtain contextualized representations of words in sentences. BERT utilizes the encoder part of transformers to encode the semantic and syntactic information of a text in embedding form, thus functioning as a language model. The pre-training procedure of BERT has two primary objectives: masked language modeling (MLM) and next sentence prediction (NSP). By using MLM, BERT can learn to predict the masked words in a sentence based on the context of the surrounding words, while NSP aims to predict whether two sentences are consecutive in a document.

The masked language modeling (MLM) objective in BERT involves randomly masking some tokens from the input and optimizing the model to predict the original vocabulary ID of the masked word based on its context. On the other hand, the BERT model has also been trained using the next sentence prediction (NSP) task to predict the text-pair representation. In BERT, a special token called '[CLS]' is added to every input sequence, whether it is a single sentence or a pair of sentences, denoted as $\langle \text{Question}, \text{Answer} \rangle$. For classification tasks, the final hidden state corresponding to the '[CLS]' token is used as an aggregated sequence representation [16].

3.2. BioBERT

BioBERT, which stands for Bidirectional Encoder Representations from Transformers for Biomedical Text Mining, is a specialized version of BERT designed for biomedical applications [16]. BioBERT is based on the contextualized language representation model of BERT, which was trained on various general and biomedical datasets, including PubMed Abstracts and PMC full-text articles. Compared to BERT and other state-of-the-art models, BioBERT has demonstrated superior performance in many NLP tasks, such as Named Entity Recognition (NER) from biomedical data, relation extraction, and question-and-answer tasks in the biomedical field.

The biomedical domain has its own specific jargon, including nouns and terms that are not present in general corpora. This creates a challenge when using general-purpose language representation models, such as BERT, for NLP tasks in the biomedical domain, as they may not perform well. To address this issue, this work utilizes BioBERT which is a biomedical domain-specific Language Representation Model based on BERT.

3.3. Relation BERT

Relation classification tasks involve predicting the semantic relationship between pairs of nominal entities. Given a sentence S and a pair of nominals e_1 and e_2 , the objective is to identify the relationship between them [52]. Numerous deep learning approaches have been proposed for relation classification, including those based on convolutional or recurrent neural networks [53,54]. Recently, pre-trained BERT models have been applied to various natural language processing tasks and have achieved state-of-the-art results in classification and SQuAD question-answering problems [55]. Although the sentence's information is crucial in classification problems, relation classification tasks also require information about the target entities.

Wu et al. [56] employed the pre-trained BERT model to improve relation classification. The authors utilized the pre-trained BERT language model and incorporated information about the target entities in the sentence to enhance the relation classification task. In general, R-BERT comprises two components: pre-trained BERT as a feature representation and additional layers that serve as a relation classifier.

The key difference between BERT and Relation BERT lies in the inputs used for their classification layers. BERT employs the final hidden state vectors of the '[CLS]' token as input, whereas R-BERT utilizes the final hidden state vectors of both the '[CLS]' token and the two entities of interest. In both cases, the '[CLS]' token is added to the beginning of each sentence and the final hidden state corresponding to the '[CLS]' token from the transformer output is used as the sentence representation for classification tasks.

3.4. Bi-Directional Long Short-Term Memory Network

LSTM networks, which stands for Long Short-Term Memory networks, are a type of recurrent neural networks (RNNs) that have the ability to learn long-term dependencies in sequence prediction tasks [57]. The architecture of LSTM networks involves three non-linear gates, namely forget gate, input gate, and output gate, which regulate the flow of information in and out of the LSTM cells. While unidirectional LSTM captures time dependencies in a single direction (either forward or backward), bidirectional LSTM preserves information in both directions (past to future and future to past) [58]. In this study, a bidirectional LSTM was utilized to extract hidden features that can access sequential data in both directions.

4. Materials and Methods

We introduce a novel method for the DDIs extraction task, leveraging a combination of R-BioBERT and BLSTM. Our proposed model is illustrated in Figure 1, and is designed to extract interactions between drugs from a sentence and classify them into specific DDIs types. In this section, we begin by outlining the dataset and the necessary steps taken

for data preprocessing. Subsequently, we describe how we incorporated R-BioBERT with BLSTM to enhance performance in the DDIs extraction task.

4.1. Datasets

In this study, we employed three datasets for DDIs extraction: the DDIs corpus from SemEval 2013 [59], TAC 2018 [60], and TAC 2019 DDIs Extraction [61].

4.1.1. SemEval 2013 DDIs Extraction

The DDIs Extraction 2011 [62] was developed for detecting drug-drug interactions in biomedical texts. Its successor, the DDIs Extraction 2013, was introduced to support additional tasks such as recognition and classification of pharmacological substances [59]. This dataset includes DrugBank with 730 documents and MEDLINE with 175 abstracts for extracting DDIs. The dataset is divided into a training set with 714 documents and a test set with 191 documents for the development and evaluation of various systems [59]. This dataset consists of four crucial DDI types: **Advice**, **Effect**, **Int**, and **Mechanism**. Drug entity pairs that are not related to these four DDI types are considered **Negative**, and there are significantly more negative DDI instances than positive ones.

- **Advice:** Advice is a type of DDI that refers to recommendations or cautions given in a document about the concurrent use of two drugs. For instance, an example of advice could be “Extreme caution should be exercised when taking alosetron and ketoconazole together”.
- **Effect:** This type in the DDIs corpus refers to the resulting effect or pharmacodynamic mechanism of interaction between two drugs. For instance, an example sentence for this type could be: “After a single administration of oxytocin, PGF2alpha caused significantly increased vasoconstriction”.
- **Int:** This refers to an interaction between drugs without providing any further information. An example of this would be “Possible interaction between atorvastatin and cyclosporine”.
- **Mechanism:** This type of DDI refers to a description of the pharmacokinetic mechanism, as in the example, “Withdrawal of rifampin decreased the warfarin requirement by 50%”.
- **Negative:** This refers to drug entity pairs that do not have any interaction. For example, “Ibogaine, but not 18-MC, decreases heart rate at high doses”.

4.1.2. TAC 2018 and TAC 2019 DDIs Extraction

The U.S. Food and Drug Administration (FDA) and the National Library of Medicine (NLM) collaborated to curate a dataset for effective deployment of drug safety information, as stated in [60]. The TAC track aims to assess the performance of NLP approaches for information extraction in DDIs, and also provides data for other tasks, such as entity extraction, relation extraction, and normalization, as mentioned in the same source. The TAC 2018 DDIs track dataset comprises 325 structured product labels (SPLs), which are divided into a training set of 22 drug labels and a test set of 57 drug labels. The types of DDIs in this dataset are classified into three categories:

- **Pharmacokinetic (PK)**
- **Pharmacodynamic (PD)**
- **Unspecified (U)**

The primary difference between TAC 2019 and TAC 2018 DDIs Extraction is that while TAC 2018 included information from structured product labels (SPLs) as well as other text types such as literature and social media, TAC 2019 only utilized SPLs. The TAC 2019 DDIs Extraction dataset comprises 406 SPLs and is divided into a training set of 211 drug labels and a test set of 81 drug labels, similar to TAC 2018. However, the types of DDIs remain consistent between the two datasets. Table 1 presents the statistics of the dataset with the official data split.

Table 1. Statistics of TAC 2018 and TAC 2019 dataset.

| DDIs | Train | | Test | |
|----------------------|----------|----------|----------|----------|
| | TAC 2018 | TAC 2019 | TAC 2018 | TAC 2019 |
| Pharmacodynamic (PD) | 47 | 553 | 335 | 292 |
| Pharmacokinetic (PK) | 60 | 494 | 296 | 118 |
| Unspecified | 62 | 665 | 440 | 202 |

4.2. Data Preprocessing

The following steps were employed for the extraction datasets of TAC 2018 and TAC 2019 DDIs:

- Firstly, instances with the same drug names in a pair were removed, as a drug cannot interact with itself. In addition, instances with only one drug in a sentence were eliminated.
- Secondly, to identify the location of two drugs in a pair, a special token $\langle e1 \rangle$ was added before and $\langle /e1 \rangle$ was added after the first drug, and $\langle e2 \rangle$ was added before and $\langle /e2 \rangle$ was added after the second drug. Unlike many other related studies, the original drug names were retained.

In the SemEval 2013 DDIs Extraction, there were significantly more negative interactions than positive interactions. This created an imbalanced class distribution problem, leading to decreased performance accuracy of the deep learning model. To address this, Zhao et al. [19] constructed a less imbalanced corpus by removing extra negative instances from the SemEval 2013 DDIs Extraction dataset using specific rules. We used a similar number of data pairs as the released code and data from this study, and applied the second step of the TAC preprocessing method to the dataset. Table 2 shows the statistics of the dataset.

Table 2. Statistics of SemEval-2013 dataset.

| DDI Samples | Training | | Test | |
|-------------|----------|----------|----------|----------|
| | Original | Filtered | Original | Filtered |
| Positive | 4020 | 3840 | 979 | 971 |
| Negative | 23,772 | 8989 | 4782 | 2084 |
| Total | 27,792 | 12,829 | 5761 | 3055 |
| Ratio | 1:5.9 | 1:2.3 | 1:4.9 | 1:2.2 |

4.3. Model Architecture

Our proposed model employs Relation BioBERT and Bidirectional Long Short-Term Memory (BLSTM) to detect and classify of DDIs. To implement the R-BioBERT model architecture, we needed to identify the location of drugs involved in the interaction and add a masked symbol before and after each target drug. As drug names do not have a fixed length, our model added $\langle e1 \rangle$ before the name of the first drug and $\langle /e1 \rangle$ at the end of the first drug. We repeated the same process for the second drug, but replaced $\langle e1 \rangle$ with $\langle e2 \rangle$. Once the location of the drugs was identified, the input was passed through BioBERT to generate the feature representation.

To demonstrate the process, consider a sentence containing target entities of “*Ganoderma lucidum extract*” and “*antibiotics*”. After inserting special tokens to indicate the target entities, the sentence is transformed as follows:

“Antimicrobial activity of $\langle e1 \rangle$ *Ganoderma lucidum extract* $\langle /e1 \rangle$ alone and in combination with some $\langle e2 \rangle$ *antibiotics* $\langle /e2 \rangle$.”

Our proposed model uses BioBERT instead of BERT, as the task of DDIs extracting involves biomedical relation extraction. Therefore, a language model trained on biomedical corpora is necessary for the model to perform accurately. Using BioBERT enhances the performance of our model by allowing it to learn from the domain-specific language and the relationships between biomedical entities.

Given a sentence S containing two entities e_1 and e_2 , we use the BioBERT model as an embedding method to obtain the final hidden state output, denoted as H . The final hidden state vectors from BioBERT for the first entity e_1 , which is composed of $token_i$ to $token_j$, are represented as H_i to H_j . Similarly, for the second entity e_2 , composed of $token_k$ to $token_m$, the final hidden state vectors are denoted as H_k to H_m .

To obtain vector representations for the two target entities, we take the average of the final hidden state vectors for each entity and apply an activation function (i.e., \tanh). Subsequently, we add a fully connected layer to each average vector to obtain the output for e_1 and e_2 , which are denoted as H'_1 and H'_2 , respectively. The calculation process is as follows in Equations (1) and (2):

$$H'_1 = W_1(\tanh(\frac{1}{j-i+1} \sum_{n=i}^j H_n)) + b_1 \quad (1)$$

$$H'_2 = W_2(\tanh(\frac{1}{k-l+1} \sum_{n=l}^k H_n)) + b_2 \quad (2)$$

$W_1 \in R^{d*d}$ and $W_2 \in R^{d*d}$ represent the vector dimension of the output layer of the BioBERT model. For more information on Relation BERT, please refer to [56].

We concatenate H'_1 and H'_2 , then apply a BLSTM layer separately to each of them. By using BERT to obtain contextualized sentence-level representations, the LSTM is better able to capture sentence semantics [63]. Recent studies have demonstrated that combining LSTM with word embedding models can yield substantial improvements in results [64]. Therefore, incorporating LSTM with BERT can lead to even better predictions, indicating a higher level of understanding of semantic meaning by the proposed model. The calculation process of BLSTM model is as follows in Equations (3)–(9):

$$\vec{f}_t = \sigma(\vec{W}_f h_{t-1} + \vec{U}_f X_t + \vec{b}_f) \quad (3)$$

$$\vec{i}_t = \sigma(\vec{W}_i h_{t-1} + \vec{U}_i X_t + \vec{b}_i) \quad (4)$$

$$c_t = \tanh(\vec{W}_c h_{t-1} + \vec{U}_c X_t + \vec{b}_c) \quad (5)$$

$$\vec{c}_t = \vec{f}_t * \vec{c}_{t-1} + \vec{i}_t * c_t \quad (6)$$

$$\vec{o}_t = \sigma(\vec{W}_o h_{t-1} + \vec{U}_o X_t + \vec{b}_o) \quad (7)$$

$$\vec{h}_t = \vec{o}_t * \tanh(\vec{c}_t) \quad (8)$$

$$\tilde{h}_t = L\tilde{S}\tilde{T}M(x_t), t \in [1, T] \quad (9)$$

Here, \vec{f}_t , \vec{i}_t , and \vec{o}_t refer to the components of the forward $L\tilde{S}\tilde{T}M$ gate, namely the forget gate, input gate, and output gate at time t , respectively. The activation function sigmoid is denoted by σ . c_t represents the input gate candidate cell. After updating at time t , \vec{c}_t represents the output of the forward memory control unit. \vec{W} and \vec{U} are the weight matrices associated with the forward class, while \vec{b} represents the offset vector of the forward class.

For the backward counterpart, denoted as $L\tilde{S}\tilde{T}M$, it follows the same structure as the forward gate in the formula. The input vector is represented by x_t . The forward \vec{h}_t is learned from x_1 to x_t at time t , and backward \tilde{h}_t is learned from x_t to x_1 at time t , and \vec{h}_t and \tilde{h}_t are concatenated to obtain the final hidden layer representation h_t as $[\vec{h}_t \oplus \tilde{h}_t]$.

Finally, The output of the BLSTM layer (h_t) is passed through a fully connected layer and a softmax layer to classify the DDIs type. The proposed model's architecture is depicted in Figure 1.

5. Experimental Evaluation

In this section, we provide details on the experimental settings and the final results obtained. We also compare our results with state-of-the-art models.

5.1. Experimental Setup

The key experimental parameters are presented in Table 3. Our experiments were carried out on a computer with a Windows operating system, equipped with a single Nvidia GeForce RTX 2070 GPU with 8GB memory. The model was implemented using the PyTorch library and the Python programming language. To prevent overfitting, early stopping was employed. Moreover, Figure 2a–c depict the training and validation loss for SemEval 2013, TAC 2018, and TAC 2019 DDI extraction datasets, respectively.

Table 3. Parameter settings.

| | |
|---------------------|--------------------|
| Batch size | 8 |
| Max sentence length | 400 |
| Adam learning rate | 2×10^{-5} |
| Number of epochs | 10 |
| Dropout rate | 0.1 |

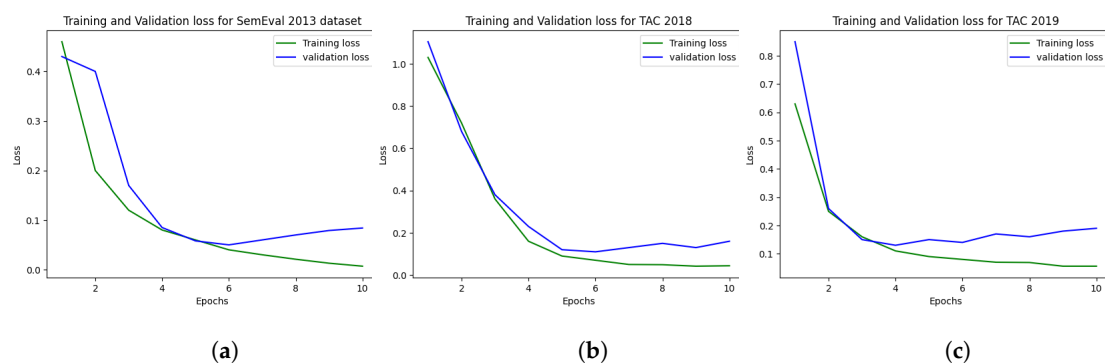


Figure 2. These figures display the training and validation loss for three datasets. (a). This plot represents the loss values for SemEval 2013. (b) This plot represents the loss values for TAC 2018. (c) This plot represents the loss values for TAC 2019.

5.2. Evaluation Metrics

The primary evaluation metrics widely used in the DDIs extraction task are F1-Macro and F1-Weighted. In this study, we assessed the performance of our R-BioBERT with BLSTM model using the Weighted-average and Macro-average F1-score on all types for SemEval 2013, and the Macro-average on all types for TAC 2018 and TAC 2019 DDIs extraction datasets.

$$Precision = \frac{TP}{(TP + FP)} \quad (10)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (11)$$

$$F1_{Score} = 2 * \left(\frac{Recall * Precision}{Recall + Precision} \right) \quad (12)$$

$$F1_{Macro} = \frac{\sum_{i=1}^5 F1_{score}}{5} \quad (13)$$

$$F1_{Weighted} = \frac{\sum_{i=1}^5 N_i * F1_i}{\sum_{i=1}^5 N_i} \quad (14)$$

N_i denotes the number of instances in class i . TP (true positive) represents the number of positive instances that are correctly classified, FP (false positive) represents the number of negative instances that are incorrectly classified as positive, and FN (false negative) represents the number of positive instances that are incorrectly classified as negative. Precision is the ratio of correctly predicted positive observations to the total predicted observations. Recall is the ratio of correctly predicted positive observations to all observations in the actual class. The $F1_{Score}$ metric is the harmonic mean of Precision and Recall metrics. The $F1_{Macro}$ score is the unweighted mean of all the per-class F1 scores, treating all classes equally. The $F1_{Weighted}$ score is calculated by taking the mean of all per-class F1 scores while considering the number of actual samples of each class.

6. Results

6.1. Results on SemEval 2013 DDIs Extraction

Table 4 displays the performance of our proposed model and state-of-the-art models on the SemEval 2013 DDI extraction task, allowing us to position our work. We calculated the Macro-average F1-score of our proposed model based on **five classes**, including the **Negative DDI** class. To facilitate comparison with related studies that employed RNNs, CNNs, or BERT, we present the results of some DDI extraction tasks, including joint AB-LSTM [20], MCCNN [33], RHCNN [34], BioBERT [48], BERT-D2 [40], EMSI-BERT [47], TP-DDI [41], BERTChem [44], IK-DDI [46], and R-BioBERT (the baseline method) [65]. As shown in Table 4, BERT-based models achieved higher overall F1-scores than RNNs and CNNs-based models.

Table 4. Model performance on SemEval 2013 DDI extraction.

| Model | F1-Score (F) | | | | | Overall Performance | |
|----------------------------------|--------------|--------------|-------------|--------------|--------------|-------------------------|--------------|
| | Negative | Mechanism | Effect | Advice | Int | F1-Score | F1-Macro |
| Joint AB-Lstm [20] | - | 72.26 | 65.46 | 80.26 | 44.11 | 69.39 | 65.52 |
| MCCNN [33] | - | 72.2 | 68.2 | 78.0 | 51.0 | 70.21 | - |
| RHCNN [34] | - | 78.3 | 73.5 | 80.5 | 58.9 | 75.5 | - |
| BioBERT [48] | - | 84.6 | 80.1 | 86 | 56.6 | 80.09 (micro-averaged) | - |
| BERT-D2 [40] | - | - | - | - | - | 81.97 | - |
| EMSI-BERT [47] | - | 86.6 | 80.07 | 86.8 | 56 | 82 (micro-averaged) | - |
| TP-DDI [41] | - | - | - | - | - | 82.4 | - |
| BERTChem [44] | - | 87 | 80 | 88 | 58 | 83 | - |
| IK-DDI [46] | - | - | - | - | - | - | 79.04 |
| R-BioBERT (Baseline) [65] | - | 97.42 | 77.80 | 87.32 | 57.31 | - | 80.89 |
| R-BioBERT with BLSTM(Our method) | 95.70 | 86.47 | 82.5 | 90.79 | 61.12 | 91.79 (weighted) | 83.32 |

Table 4 presents the performance comparison of various DDIs extraction models. Among the CNN and LSTM-based models, the Joint AB-Lstm model had the lowest overall F1-score (69.39%), and its best F1-score is reported for Mechanism (72.26%), while the worst performance belonged to Int (44.11%). In comparison, MCCNN achieved a better overall F1-score of 70.21% than the Joint AB-Lstm, but was inferior to RHCNN (75.5%).

Regarding BERT-based models, BioBERT had the lowest overall F1-Micro (80.09%). On the other hand, BERT-D2 had a better overall F1-score (81.97%) than BioBERT, although it was inferior to EMSI-BERT F1-Micro (82%). Additionally, TP-DDI achieved an F1-micro of 82.04%, while BERTchem achieved a higher F1-Micro of 83%. Furthermore, the IK-DDI obtained an overall F1-Macro of 79.04%. However, the R-BioBERT, which was the baseline model, outperformed the IK-DDI model with an F1-Macro of 80.89%.

Finally, our proposed model, R-BioBERT with BLSTM, achieved the best overall F1-weighted (91.79%) and F1-Macro (83.32%) compared to the state-of-the-art models.

Table 4 highlights that R-BioBERT achieved the highest F1-score (97.42%) while MCCNN reported the lowest F1-score (72.2%) in Mechanism. However, our model (R-BioBERT with BLSTM) achieved F1-score of 86.47% in Mechanism which was lower than the performance of R-bioBERT model in the Mechanism class. The observed significant difference can likely be attributed to several factors. Firstly, the baseline model (R-BioBERT) did not account for negative types of interactions and only considered four classes of DDIs. In contrast, our work considered five classes, namely Mechanism, Effect, Advice, Int, and Negative. Additionally, a comparison in Table 4 reveals that our model outperformed both the baseline model and other state-of-the-art models in classifying the four DDIs types, including Negative (95.70%), Effect (82.5%), Advice (90.79%), and Int (61.12%). This comprehensive evaluation demonstrates that our model excels across these classes, highlighting its superiority and establishing its reliability.

The worst F1-scores for Effect and Int belonged to the Joint AB-Lstm. Additionally, in Advice, the highest F1-score belonged to the R-BioBERT with BLSTM model, while MCCNN had the worst F1-score. The highest F1-scores are indicated in bold in Table 4.

Furthermore, R-BioBERT with BLSTM exhibited the best overall performance with the highest F1-score (91.79%) and F1-Macro (83.32%). Our proposed model significantly outperforms previous CNN and LSTM-based models, especially the baseline method [65]. The F1-Macro of R-BioBERT with BLSTM is 83.32, which is substantially better than the previous best solution on the SemEval 2013 dataset. Our model also achieves better performance in the Negative, Effect, and Int classes compared to previous research [65]. Notably, the Int class exhibits limited performance across all models, possibly due to insufficient training data.

6.2. Results on TAC 2018

Table 5 presents the evaluation results of our proposed model and the state-of-the-art models on the TAC 2018 DDIs extraction task. In this task, the F1-macro of our proposed model is calculated based on three classes.

Table 5. Model performance on TAC 2018 DDI extraction.

| Model | F1-Score (F) | | | Overall Performance |
|-----------------------------------|--------------|------|------|----------------------|
| | Unspecified | PK | PD | F1-Score |
| [66] | - | - | - | 40.90 |
| [67] | - | - | - | 56.98 |
| R-BioBERT with BLSTM (Our method) | 64 | 58.8 | 58.8 | 60.53 (Macro) |

The results in Table 5 indicate that Tang et al. [66] achieved the lowest F1-score (40.90%), even though they used not only the TAC 2018 dataset but also NLM-180 and HS datasets. On the other hand, the model proposed in [67] achieved a better F1-score (56.98%) than Tang et al. [66]. However, our proposed model, R-BioBERT with BLSTM, outperforms all the state-of-the-art models with an F1-Macro score of 62.64%.

The relatively small size of the TAC 2018 dataset compared to SemEval 2013 and TAC 2019 may be one reason for the low F1-scores in this task. Therefore, achieving a higher F1-score on the TAC 2018 dataset is challenging and our proposed model demonstrates superior performance in this regard.

6.3. Results on TAC 2019

Table 6 presents the performance of our proposed model and the state-of-the-art models on the TAC 2019 DDIs extraction task. We can observe that the model proposed by Mahajan et al. [68] had the lowest F1-score (40.39%), while UTDHLTRI 2 [61] achieved a slightly better F1-score (49.2%) than IBMResearch 1 [61] (50.1%). Our proposed model, R-BioBERT with BLSTM, outperforms the state-of-the-art models by a large margin, achieving the highest F1-score of 80.26%.

Table 6. Model performance on TAC 2019 DDI extraction.

| Model | F1-Score (F) | | | Overall Performance |
|-----------------------------------|--------------|------|------|----------------------|
| | Unspecified | PK | PD | F1-Score |
| [68] | 48.3 | 63.2 | 43.4 | 40.39 |
| UTDHLTRI 2 [61] | - | - | - | 49.2 |
| IBMResearch 1 [61] | - | - | - | 50.1 |
| R-BioBERT with BLSTM (Our method) | 76.2 | 81.2 | 83.3 | 80.23 (Macro) |

The superior performance of our proposed model can be attributed to the use of the BioBERT language model, which is specifically pre-trained on biomedical texts, and the BLSTM layer, which is effective in capturing long-range dependencies in text data. The TAC 2019 dataset is more challenging than TAC 2018, and our model's excellent performance indicates its effectiveness in extracting DDIs from complex biomedical texts.

In conclusion, our proposed model achieves state-of-the-art performance on both TAC 2018 and TAC 2019 DDIs extraction tasks, demonstrating its effectiveness in biomedical text mining applications.

7. Conclusions

The present study introduces a novel method for DDIs extraction by integrating Relation-BioBERT and BLSTM. Our experimental results demonstrate that the proposed approach outperforms existing models for DDIs extraction on Semeval 2013, TAC 2018, and TAC 2019 datasets. Specifically, our model achieves an F1-Macro score of 83.32% on SemEval, 80.23% on TAC 2018, and 60.53% on TAC 2018 for DDI extraction.

There are several potential directions for improving and extending our approach. Firstly, our model's performance in classifying mechanism interactions is not yet satisfactory, and we plan to explore strategies to improve performance with limited training data. Additionally, we aim to apply our proposed method to other relation extraction tasks beyond DDI extraction.

Author Contributions: Methodology, M.K.; Writing—original draft, M.K.; Writing—review & editing, A.H.; Supervision, A.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The code will be available upon request to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Miranda, V.; Fede, A.; Nobuo, M.; Ayres, V.; Giglio, A.; Miranda, M.; Riechelmann, R.P. Adverse drug reactions and drug interactions as causes of hospital admission in oncology. *J. Pain Symptom Manag.* **2011**, *42*, 342–353. [\[CrossRef\]](#)
2. Lazarou, J.; Pomeranz, B.H.; Corey, P.N. Incidence of adverse drug reactions in hospitalized patients: A meta-analysis of prospective studies. *JAMA* **1998**, *279*, 1200–1205. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Becker, M.L.; Kallewaard, M.; Caspers, P.W.; Visser, L.E.; Leufkens, H.G.; Stricker, B.H. Hospitalisations and emergency department visits due to drug–drug interactions: A literature review. *Pharmacoepidemiol. Drug Saf.* **2007**, *16*, 641–651. [\[CrossRef\]](#)
4. Businaro, R. Why we need an efficient and careful pharmacovigilance? *J. Pharmacovigil.* **2013**. [\[CrossRef\]](#)
5. Hohl, C.M.; Dankoff, J.; Colacone, A.; Afilalo, M. Polypharmacy, adverse drug-related events, and potential adverse drug interactions in elderly patients presenting to an emergency department. *Ann. Emerg. Med.* **2001**, *38*, 666–671. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Paczynski, R.P.; Alexander, G.C.; Chinchilli, V.M.; Kruszewski, S.P. Quality of evidence in drug compendia supporting off-label use of typical and atypical antipsychotic medications. *Int. J. Risk Saf. Med.* **2012**, *24*, 137–146. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Rodríguez-Terol, A.; Caraballo, M.; Palma, D.; Santos-Ramos, B.; Molina, T.; Desongles, T.; Aguilar, A. Quality of interaction database management systems. *Farm. Hosp. (Engl. Ed.)* **2009**, *33*, 134–146. [\[CrossRef\]](#)
8. Ammar, W.; Groeneveld, D.; Bhagavatula, C.; Beltagy, I.; Crawford, M.; Downey, D.; Dunkelberger, J.; Elgohary, A.; Feldman, S.; Ha, V.; et al. Construction of the literature graph in semantic scholar. *arXiv* **2018**, arXiv:1805.02262.
9. Asada, M.; Miwa, M.; Sasaki, Y. Enhancing drug–drug interaction extraction from texts by molecular structure information. *arXiv* **2018**, arXiv:1805.05593.
10. Beltagy, I.; Lo, K.; Cohan, A. SciBERT: A pretrained language model for scientific text. *arXiv* **2019**, arXiv:1903.10676.

11. Zhang, T.; Leng, J.; Liu, Y. Deep learning for drug–drug interaction extraction from the literature: A review. *Brief. Bioinform.* **2020**, *21*, 1609–1627. [\[CrossRef\]](#)
12. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#)
13. Lauriola, I.; Lavelli, A.; Aiolli, F. An introduction to deep learning in natural language processing: Models, techniques, and tools. *Neurocomputing* **2022**, *470*, 443–456. [\[CrossRef\]](#)
14. Pennington, J.; Socher, R.; Manning, C.D. Glove: Global vectors for word representation. In Proceedings of the 2014 Conference On Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1532–1543.
15. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. *arXiv* **2013**, arXiv:1301.3781.
16. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
17. Shen, S.; Liu, J.; Lin, L.; Huang, Y.; Zhang, L.; Liu, C.; Feng, Y.; Wang, D. SsciBERT: A pre-trained language model for social science texts. *Scientometrics* **2023**, *128*, 1241–1263. [\[CrossRef\]](#)
18. Hong, L.; Lin, J.; Li, S.; Wan, F.; Yang, H.; Jiang, T.; Zhao, D.; Zeng, J. A novel machine learning framework for automated biomedical relation extraction from large-scale literature repositories. *Nat. Mach. Intell.* **2020**, *2*, 347–355. [\[CrossRef\]](#)
19. Zhao, Z.; Yang, Z.; Luo, L.; Lin, H.; Wang, J. Drug drug interaction extraction from biomedical literature using syntax convolutional neural network. *Bioinformatics* **2016**, *32*, 3444–3453. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Sahu, S.K.; Anand, A. Drug-drug interaction extraction from biomedical texts using long short-term memory network. *J. Biomed. Inform.* **2018**, *86*, 15–24. [\[CrossRef\]](#)
21. Williams, R.J.; Zipser, D. A learning algorithm for continually running fully recurrent neural networks. *Neural Comput.* **1989**, *1*, 270–280. [\[CrossRef\]](#)
22. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [\[CrossRef\]](#)
23. Jiang, Y.; Chen, L.; Zhang, H.; Xiao, X. Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module. *PLoS ONE* **2019**, *14*, e0214587. [\[CrossRef\]](#)
24. Dai, X.; Chen, Y.; Xiao, B.; Chen, D.; Liu, M.; Yuan, L.; Zhang, L. Dynamic head: Unifying object detection heads with attentions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 7373–7382.
25. Behzadi, M.M.; Ilies, H.T. Real-time topology optimization in 3d via deep transfer learning. *Comput.-Aided Des.* **2021**, *135*, 103014. [\[CrossRef\]](#)
26. Basiri, M.E.; Nemati, S.; Abdar, M.; Cambria, E.; Acharya, U.R. ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis. *Future Gener. Comput. Syst.* **2021**, *115*, 279–294. [\[CrossRef\]](#)
27. Shen, Y.; He, X.; Gao, J.; Deng, L.; Mesnil, G. Learning semantic representations using convolutional neural networks for web search. In Proceedings of the 23rd International Conference on World Wide Web, Seoul, Republic of Korea, 7–11 April 2014; pp. 373–374.
28. Yih, W.T.; He, X.; Meek, C. Semantic parsing for single-relation question answering. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Baltimore, MD, USA, 22–27 June 2014; pp. 643–648.
29. Liu, S.; Tang, B.; Chen, Q.; Wang, X. Drug-drug interaction extraction via convolutional neural networks. *Comput. Math. Methods Med.* **2016**, *2016*, 6918381. [\[CrossRef\]](#)
30. Asada, M.; Miwa, M.; Sasaki, Y. Extracting Drug-Drug Interactions with Attention CNNs. In *BioNLP 2017*; Association for Computational Linguistics: Vancouver, BC, Canada, 2017; pp. 9–18. [\[CrossRef\]](#)
31. Dewi, I.N.; Dong, S.; Hu, J. Drug-drug interaction relation extraction with deep convolutional neural networks. In Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Kansas City, MO, USA, 13–16 November 2017; IEEE Computer Society: Washington, DC, USA, 2017; pp. 1795–1802. [\[CrossRef\]](#)
32. Sun, X.; Ma, L.; Du, X.; Feng, J.; Dong, K. Deep convolution neural networks for drug-drug interaction extraction. In Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine (Bibm), Madrid, Spain, 3–6 December 2018; IEEE: New York, NY, USA, 2018; pp. 1662–1668. [\[CrossRef\]](#)
33. Quan, C.; Hua, L.; Sun, X.; Bai, W. Multichannel convolutional neural network for biological relation extraction. *BioMed Res. Int.* **2016**, *2016*, 1850404. [\[CrossRef\]](#)
34. Sun, X.; Dong, K.; Ma, L.; Sutcliffe, R.; He, F.; Chen, S.; Feng, J. Drug-drug interaction extraction via recurrent hybrid convolutional neural networks with an improved focal loss. *Entropy* **2019**, *21*, 37. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Kavuluru, R.; Rios, A.; Tran, T. Extracting drug-drug interactions with word and character-level recurrent neural networks. In Proceedings of the 2017 IEEE International Conference on Healthcare Informatics (ICHI), Park City, UT, USA, 23–26 August 2017; IEEE: New York, NY, USA, 2017; pp. 5–12. [\[CrossRef\]](#)
36. Huang, D.; Jiang, Z.; Zou, L.; Li, L. Drug-drug interaction extraction from biomedical literature using support vector machine and long short term memory networks. *Inf. Sci.* **2017**, *415*, 100–109. [\[CrossRef\]](#)
37. Yi, Z.; Li, S.; Yu, J.; Tan, Y.; Wu, Q.; Yuan, H.; Wang, T. Drug-drug interaction extraction via recurrent neural network with multiple attention layers. In *Advanced Data Mining and Applications, Proceedings of the 13th International Conference, ADMA 2017, Singapore, 5–6 November 2017*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 554–566.

38. Zhou, D.; Miao, L.; He, Y. Position-aware deep multi-task learning for drug–drug interaction extraction. *Artif. Intell. Med.* **2018**, *87*, 1–8. [CrossRef] [PubMed]
39. Peng, Y.; Yan, S.; Lu, Z. Transfer learning in biomedical natural language processing: An evaluation of BERT and ELMo on ten benchmarking datasets. *arXiv* **2019**, arXiv:1906.05474.
40. Datta, T.T.; Shill, P.C.; Al Nazi, Z. BERT-D2: Drug-Drug Interaction Extraction using BERT. In Proceedings of the 2022 International Conference for Advancement in Technology (ICONAT), Goa, India, 21–22 January 2022; IEEE: New York, NY, USA, 2022; pp. 1–6.
41. Zaikis, D.; Vlahavas, I. TP-DDI: Transformer-based pipeline for the extraction of Drug-Drug Interactions. *Artif. Intell. Med.* **2021**, *119*, 102153. [CrossRef] [PubMed]
42. Lee, J.; Yoon, W.; Kim, S.; Kim, D.; Kim, S.; So, C.H.; Kang, J. BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* **2020**, *36*, 1234–1240. [CrossRef] [PubMed]
43. Rasmy, L.; Xiang, Y.; Xie, Z.; Tao, C.; Zhi, D. Med-BERT: Pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *NPJ Digit. Med.* **2021**, *4*, 1–13. [CrossRef] [PubMed]
44. Mondal, I. Bertchem-ddi: Improved drug-drug interaction prediction from text using chemical structure information. *arXiv* **2020**, arXiv:2012.11599.
45. Asada, M.; Miwa, M.; Sasaki, Y. Using drug descriptions and molecular structures for drug–drug interaction extraction from literature. *Bioinformatics* **2021**, *37*, 1739–1746. [CrossRef]
46. Dou, M.; Ding, J.; Chen, G.; Duan, J.; Guo, F.; Tang, J. IK-DDI: A novel framework based on instance position embedding and key external text for DDI extraction. *Brief. Bioinform.* **2023**, *24*, bbad099. [CrossRef]
47. Huang, Z.; An, N.; Liu, J.; Ren, F. EMSI-BERT: Asymmetrical Entity-Mask Strategy and Symbol-Insert Structure for Drug–Drug Interaction Extraction Based on BERT. *Symmetry* **2023**, *15*, 398. [CrossRef]
48. Zhu, Y.; Li, L.; Lu, H.; Zhou, A.; Qin, X. Extracting drug-drug interactions from texts with BioBERT and multiple entity-aware attentions. *J. Biomed. Inform.* **2020**, *106*, 103451. [CrossRef]
49. Peters, M.E.; Neumann, M.; Iyyer, M.; Gardner, M.; Clark, C.; Lee, K.; Zettlemoyer, L. Deep Contextualized Word Representations. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), New Orleans, LA, USA, 1–6 June 2018; pp. 2227–2237. [CrossRef]
50. McCann, B.; Bradbury, J.; Xiong, C.; Socher, R. Learned in translation: Contextualized word vectors. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. [CrossRef]
51. Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. Improving Language Understanding by Generative Pre-Training. 2018. Available online: https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf (accessed on 16 May 2023).
52. Hendrickx, I.; Kim, S.N.; Kozareva, Z.; Nakov, P.; Séaghdha, D.O.; Padó, S.; Pennacchiotti, M.; Romano, L.; Szpakowicz, S. Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals. *arXiv* **2019**, arXiv:1911.10422.
53. Nogueira dos Santos, C.; Xiang, B.; Zhou, B. Classifying relations by ranking with convolutional neural networks. *arXiv* **2015**, arXiv:1504.06580.
54. Lee, J.; Seo, S.; Choi, Y.S. Semantic relation classification via bidirectional lstm networks with entity-aware attention using latent entity typing. *Symmetry* **2019**, *11*, 785. [CrossRef]
55. Rajpurkar, P.; Zhang, J.; Lopyrev, K.; Liang, P. Squad: 100,000+ questions for machine comprehension of text. *arXiv* **2016**, arXiv:1606.05250.
56. Wu, S.; He, Y. Enriching pre-trained language model with entity information for relation classification. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, Beijing, China, 3–7 November 2019; pp. 2361–2364.
57. Wu, Y.; Yuan, M.; Dong, S.; Lin, L.; Liu, Y. Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. *Neurocomputing* **2018**, *275*, 167–179. [CrossRef]
58. Zhou, P.; Qi, Z.; Zheng, S.; Xu, J.; Bao, H.; Xu, B. Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling. *arXiv* **2016**, arXiv:1611.06639.
59. Herrero-Zazo, M.; Segura-Bedmar, I.; Martínez, P.; Declerck, T. The DDI corpus: An annotated corpus with pharmacological substances and drug–drug interactions. *J. Biomed. Inform.* **2013**, *46*, 914–920. [CrossRef]
60. Demner-Fushman, D.; Fung, K.W.; Do, P.; Boyce, R.D.; Goodwin, T.R. Overview of the TAC 2018 Drug-Drug Interaction Extraction from Drug Labels Track. In Proceedings of the TAC, Gaithersburg, MD, USA, 13–14 November 2018.
61. Goodwin, T.R.; Demner-Fushman, D.; Fung, K.W.; Do, P. Overview of the TAC 2019 Track on Drug-Drug Interaction Extraction from Drug Labels. In Proceedings of the TAC, Gaithersburg, MD, USA, 12–13 November 2019.
62. Segura-Bedmar, I.; Martínez Fernández, P.; Sánchez Cisneros, D. *The 1st DDIExtraction-2011 Challenge Task: Extraction of Drug-Drug Interactions from Biomedical Texts*; In Proceedings of the 1st Challenge Task on Drug-Drug Interaction Extraction, Huelva, Spain, 11 September 2011.
63. Rai, N.; Kumar, D.; Kaushik, N.; Raj, C.; Ali, A. Fake News Classification using transformer based enhanced LSTM and BERT. *Int. J. Cogn. Comput. Eng.* **2022**, *3*, 98–105. [CrossRef]
64. Deepak, S.; Chitturi, B. Deep neural approach to Fake-News identification. *Procedia Comput. Sci.* **2020**, *167*, 2236–2243.
65. Nguyen, D.P.; Ho, T.B. Drug-drug interaction extraction from biomedical texts via relation BERT. In Proceedings of the 2020 RIVF International Conference on Computing and Communication Technologies (RIVF), Ho Chi Minh City, Vietnam, 14–15 October 2020; IEEE: New York, NY, USA, 2020; pp. 1–7. [CrossRef]

66. Tang, S.; Zhang, Q.; Zheng, T.; Zhou, M.; Chen, Z.; Shen, L.; Ren, X.; Zhuang, Y.; Pu, S.; Wu, F. Two step joint model for drug drug interaction extraction. *arXiv* **2020**, arXiv:2008.12704.
67. Tran, T.; Kavuluru, R.; Kilicoglu, H. A multi-task learning framework for extracting drugs and their interactions from drug labels. *arXiv* **2019**. arXiv:1905.07464.
68. Mahajan, D.; Poddar, A.; Lin, Y.T. A hybrid model for drug-drug interaction extraction from structured product labeling documents. In Proceedings of the TAC, Gaithersburg, MD, USA, 12–13 November 2019.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.