

Received 5 February 2025, accepted 13 March 2025, date of publication 18 March 2025, date of current version 25 March 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3552239

RESEARCH ARTICLE

Innovative Tailored Semantic Embedding and Machine Learning for Precise Prediction of Drug-Drug Interaction Seriousness

AYMAN MOHAMED MOSTAFA^{ID1}, (Member, IEEE), ALAA S. ALAERJAN^{ID2},
HISHAM ALLAHEM¹, BADER ALDUGHAYFIQ^{ID1}, MESHRIF ALRUILY^{ID2},
ALSHAIMAA A. TANTAWY³, AND MOHAMED EZZ^{ID2}, (Member, IEEE)

¹Information Systems Department, College of Computer and Information Sciences, Jouf University, Sakaka 72388, Saudi Arabia

²Department of Computer Science, College of Computer and Information Sciences, Jouf University, Sakaka 72388, Saudi Arabia

³Department of Information Systems, Faculty of Computers and Informatics, Zagazig University, Zagazig 44519, Egypt

Corresponding author: Ayman Mohamed Mostafa (amhassane@ju.edu.sa)

This work was supported by the Deanship of Graduate Studies and Scientific Research at Jouf University under Grant DGSSR-2023-02-02472.

ABSTRACT This study explores applying advanced machine-learning strategies, particularly improved semantic vectors, to predict the severity of drug-drug interactions (DDIs), a crucial element in pharmacovigilance. Based on the Adverse Event Reporting System (FAERS), our study aims to analyze the combination of advanced embedding techniques with state-of-the-art machine learning (ML) algorithms to identify and quantify DDI severity. The CatBoost Classifier is the center of our analysis, as it has emerged as the most effective model in the examined trials. We improved the performance by increasing the BioWordVec Indication Substance embedding specificity, a new creation constructed through transfer learning methodologies employed on the BioWordVec model. This approach employs not only the names of the drugs but also the indications for the drugs and the active substances, forming a highly semantic network capable of capturing multiple relations between drugs. Applying BioWordVec Indication Substance embedding combined with the CatBoost Classifier, especially using the contact-vectors method, provided the best F1 score of 73.32% and an ROC AUC score of 84%. The results imply that this method effectively models and predicts severe consequences of DDIs using deep learning that comprehensively covers pharmacological and clinical aspects. Based on our results, we suggest incorporating semantic embedding and ML into the pharmacovigilance processes to improve the predictive potential of DDI evaluations. Thus, by enhancing the body of knowledge related to the analytical methods of assessing drug interactions, the present study substantially enhances the quality of clinical decision-making and patient protection. The novel embedding marks a significant step forward in the methodology, providing a more solid tool for the fine-grained dissection of the complexities needed in modern medicine, where multiple drug therapies are now the norm.

INDEX TERMS Transfer learning, drug-drug interaction (DDI), CatBoost classifier, BioWordVec, semantic embedding.

I. INTRODUCTION

The growth of the number of different medications and the complexity of the pharmaceutical treatments have raised

The associate editor coordinating the review of this manuscript and approving it for publication was Mu-Yen Chen^{ID}.

the risks of drug-drug interactions (DDIs) and adverse drug reactions (ADRs) that have a significant impact on the safety of patients. These interactions are essential to understand and predict in clinical medicine to avoid the adverse effects of polypharmacy. FAERS by the Food and Drug Administration collects a vast database of adverse effects

associated with the use of drugs, which is reported voluntarily. The abundance of data available in FAERS provides an excellent opportunity to investigate and forecast the outcomes of DDIs employing sophisticated analytics [1]. The developments in ML and natural language processing have offered approaches that can be used to analyze the extensive databases in pharmacovigilance. Specifically, text vectorization and semantic embedding methods have been immensely helpful in converting drug data into structures that can be analyzed computationally [2], [3]. This study employs these technological enhancements to develop a system that works on the FAERS dataset to enhance the prediction of ADRs due to DDIs. Through advanced mathematical models, this study seeks to improve the comprehension of multi-drug interactions to reduce drug-related dangers and provide better therapeutic decision-making.

The BioWordVec was selected in this study due to its capability of delivering domain-specific embedding designed for biomedical and pharmacological domains. In contrast to general embedding, BioWordVec has been trained on the biomedical text and, therefore, is appropriate for drug-drug interaction evaluation. Furthermore, fine-tuning BERT-based models [4], [5] or any other Transformer-based biomedical embedding like BioBERT may be possible. However, they consume much more computational power than the embedding models described here. This research highlights the following questions:

Q1: How can text vectorization and ML be utilized to predict the seriousness of drug interactions based on adverse event reports in the FAERS database?

- This question explores the application of count vectorization and ML algorithms to determine the potential severity of DDIs. We will investigate whether these computational techniques can effectively predict serious outcomes when two or more drugs are taken concurrently.

Q2: What role does semantic embedding from pre-trained biomedical models play in enhancing the predictive accuracy of drug interaction seriousness?

- We aim is to explore the potential of BioWordVec, a pre-trained biomedical word embedding model, to capture semantic relationships between drug names. This question will delve into whether these embedding can provide a deeper understanding of DDIs and improve prediction accuracy compared to traditional vectorization methods.

Q3: Can transfer learning be effectively applied to adapt existing word embedding to better predict DDIs in the context of specific patient demographics and drug characteristics?

- This question considers the adaptability of pre-trained embedding through transfer learning, focusing on customizing these models to incorporate additional drug-related features such as indications and active substances. We will explore the potential improvements in predictive performance when these adapted embedding are used to assess the seriousness of DDIs.

This study aims to provide a comprehensive analysis of DDIs using innovative data processing techniques and ML models, contributing significantly to the fields of pharmacovigilance and clinical informatics. By answering these questions, we hope to advance the methodologies used in drug safety research and improve the outcomes of drug therapy in clinical settings.

II. RELATED WORKS

In recent decades, there has been a good deal of progress in applying computational approaches for the analysis of DDIs, with the aid of a range of datasets and tools. In this paper, we have presented and discussed several research papers which have shaped the field of prediction and analysis of DDIs. As described in [6], the study proposes to increase the early identification of DDIs by creating and testing a technique to identify DDI alerts from the EHRs. The method used here emphasizes signal detection through clinical notes. This work shows how EHRs may be used to identify adverse drug effects because they replicate previous traditional pharmacovigilance and compare their results to clinical notes and prior research using the FAERS and a benchmark set of established DDIs. The authors of [7] intend to establish and characterize ADRs associated with counterfeit drugs using data from the pharmacovigilance domain. When aggregated from the FAERS dataset on safety concerns, the work considers cases of fakes or goods that have been modified in one way or another. Furthermore, the authors of [8] lay down systematic screening for potential ADEs associated with DDs that have PK interaction mechanisms. This is done with the help of a comprehensive surveillance database and a translational informatics approach: CYP isoforms and ADE reports from FAERS.

Applying ML and deep learning algorithms to predict DDI is considered one of the main parameters for enhancing the quality of clinical decision-making. As presented in [9], using ML techniques, text mining systems are presented and evaluated for drug name recognition, DDI extraction, and DDI classification. These systems were tested on benchmarks from the DDI-Extraction shared tasks of 2011 and 2013 to advance pharmacovigilance by improving the detection and classification of drug-drug interactions. While the method contributes to data extraction techniques, it falls short in embedding complex drug interaction data. The authors of [10] reviewed the progress of ML methods in predicting unknown DDIs, highlighting the literature-based approach that combines DDI extraction and prediction methods. The paper introduces common databases, describes various prediction models, summarizes their advantages and disadvantages, and discusses the challenges and prospects in the field. Furthermore, the authors of [11] introduced an interactive biomedical text mining framework for identifying drug names from unstructured biomedical text. This framework enables real-time extraction of drug entities from scientific abstracts on PubMed and benchmark datasets, using

advanced ML techniques for preprocessing and recognition. As presented in [12], a novel approach for detecting and classifying DDI is proposed by combining Relation BioBERT (R-BioBERT) with Bidirectional Long Short-Term Memory (BLSTM) models. This method aims to enhance the accuracy of DDI predictions and identify the specific types of interactions between drugs.

The authors of [13] discuss a study that used SMILES codes to create features for over 5,000 medications from DrugBank and developed a deep neural network model to predict 80 drug-drug interactions (DDIs). The predictions were then applied to analyze inflammatory bowel disease (IBD) medication regimens, sparking discussion on potential treatment strategies. Recently, adverse drug-drug interaction (ADDI) can be measured as a significant challenge in pharmacological production, as it has become a leading cause of illness and death. The authors of [14] describe a neural network method to predict novel, unknown ADDIs using different drug-related data, such as substructure, target, side effects, off-label side effects, route, transporter, and indication data.

The authors of [15] give a multimodal deep learning framework called DDIMDL. The work of Zhang et al. incorporated deep learning with several pharmacological features. In light of four drug features, including chemical substructures, targets, enzymes and pathways, DDIMDL successfully built four kinds of DNN-based sub-models. Following the work done by Mei and Zhang [16], some attempt is made to look at drug-drug interactions by analyzing the interactions between the genes affected by two drugs. It is possible to define several statistical characteristics that describe the strengths, efficiency, and breadth of the action of two drugs in human cells' PPI networks and signaling pathways. An Interaction Prediction Graph Attention Network (IPGAT) paradigm is proposed by Wang et al. [17]. It comprised two modules: the prediction module and the embedding module. The embedding module took its cues from the Graph Embedding and Graph Attention Networks to extract features with high-order neighborhoods from graph-structured data. Another model for predicting the cold start of single-type and multiple-type DDIs was introduced by Liu et al. [18] under the name CSMDDI. In the cold start scenario, CSMDDI did not only predict whether or not two medications elicited pharmacological reactions but also the type of reaction that the medications would cause. To evaluate the performance, they used CSMDDI with the competing multi-type DDI prediction methods. A comprehensive analysis of the aid of AI and ML in the early detection of adverse drug reactions (ADRs) and toxicity was provided in [19]. They incorporate a variety of methodologies, including data mining and deep learning, and then provide a list of significant databases, modelling algorithms, and software that can be used to model and predict a number of ADRs and toxicity.

By improving the specificity of BioWordVec embedding, your method addresses some of the key challenges

highlighted in this review, such as the need for more precise predictive models. Based on heterogeneous signed networks, the authors of [20] propose Adaptive Dual Graph Contrastive Learning (ADGCL) as a novel approach for predicting Adverse Drug Reactions (ADRs). ADGCL first explicitly models both positive and negative DDIs via a heterogeneous signed network in order to generate semantically rich drug feature representations. It is still crucial to investigate a strong projection between drug attributes and their adverse interactions to uncover their adverse correlation's non-linear characteristics for accurate ADDI forecast, given the high dimensionality and extreme sparsity of the hand-designed drug attributes [21]. In [22], a different Siamese-like architecture with deep convolutional transform learning for two processing channel networks was developed. Fused and channel-wise representations were also learned from the transformation across them. A decision forest was given the final representation as the last layer to generate the final predictions.

Using GNNs, or graph neural networks, DDI was developed as a hypergraph by Nguyen et al. [23] with two nodes for medications and one for a label for each hyperedge. Then, using a novel "central-smoothing" formulation, they introduced CentSmoothie, a hypergraph neural network (HGNN) that learns representations of combined nodes and labels. Furthermore, the authors of [24] assessed ChatGPT's ability to predict and explain drug-drug interactions (DDIs) by querying it with 40 DDI pairs and evaluating its responses. The results showed that while ChatGPT can provide helpful information about DDIs. To efficiently and automatically develop the GNN architecture for drug-drug interaction prediction without manual involvement, the authors of [25] automate the development of GNN architectures for predicting drug-drug interactions (DDIs). This paper automates the development of GNN architectures for DDI prediction, offering a systematic approach to architecture search.

To capture the multimodal properties of pharmaceuticals, the authors of [26] developed the MSKG-DDI. This two-component framework combines the Drug Chemical Structure Graph-based and Drug Knowledge Graph-based components. The complementarity among multimodal representations of medications was then investigated using a multimodal fusion neural layer. In the DDI datasets that are currently available, there are comparatively few positive examples, according to the [27] assessment. It is challenging for deep learning (DL) algorithms to extract enough feature information from text data directly. Consequently, current deep learning models primarily depend on various feature replenishment techniques to obtain adequate feature information from various data formats. Different research methods also focused on vectorization and feature engineering in DDI, as presented in [28] and [29]. The authors of [28] underline the transition from manual feature engineering to the deep learning approach that involves word embedding and distance embedding in the framework of the stacked Bi-LSTM-CNN

model, where the authors of [29] compare different methods of vectorization with the more recent word embedding techniques such as Word2Vec and GloVe to assess their efficiency in opinion mining from drug reviews. The paper states that the research aims to compare the traditional count vectorization methods with the modern embedding methods to determine the effective model for analyzing patients' emotions toward drugs. As presented in [30], the authors improve the quality of word representations utilized in biomedical natural language processing (BioNLP). In particular, the abstract describes BioWordVec, a set of biomedical word vectors/embedding trained on the sub-word information from the biomedical text without labels and the structured domain-specific information from the Medical Subject Headings (MeSH). Finally, transfer learning applications can also be applied in pharmacovigilance, as presented in [31], where the purpose of the study is to describe how pharmacovigilance differs from other domains and to pinpoint areas where ML can be used to advance pharmacovigilance. Table 1 explores the major performance of recent DDI methodologies to explain the applied algorithms and methods.

III. METHODOLOGY

Our study's methodology harnesses the power of ML and natural language processing to predict DDIs and assess their seriousness based on data from the FAERS database. As presented in Figure 1, a multifaceted approach combines text vectorization, semantic embedding, and advanced ML techniques. Figure 1 illustrates our proposed approach for predicting the severity of drug-drug interactions (DDIs). The methodology follows a structured pipeline that combines text vectorization, semantic embedding, and machine learning techniques. The process starts with data preprocessing, where we extract drug pairs from the FAERS dataset, remove duplicates, handle missing values, and categorize interactions as serious or non-serious.

To improve model performance, we employed advanced semantic embedding techniques such as BioWordVec, enabling the model to capture the contextual relationships between drug names, active substances, and indications. However, since standard BioWordVec embedding had only 40% coverage of the FAERS dataset, we refined them into BioWordVecIndicationSubstance embedding, which integrate both drug indications and active substances to create a more comprehensive feature representation. This enhancement significantly improved embedding coverage to 80%, allowing the model to better understand pharmacological relationships and interactions. We conducted different vector combination techniques to achieve the best possible drug pair representation. The evaluation process revealed concatenation as the most successful method among multiple approaches which included vector difference and distance-based methods and concatenation.

The complete semantic and pharmacological characteristics of both drugs remain intact in the final representation through this method. The model utilizes an enhanced

embedding approach together with an optimal vector combination to reach higher predictive accuracy when evaluating DDI severity.

After vectorizing the drug pairs, we train and evaluate several ML models, including Logistic Regression, K-Nearest Neighbors, Naïve Bayes, XGBoost, LightGBM, Extra Trees, and CatBoost. CatBoost is unique in efficiently handling high-dimensional categorical data while minimizing the need for extensive feature preprocessing.

The figure shows also how our methodology highlights the need for incorporating domain-specific semantic embedding with advanced classification models to increase the prediction of DDI severity.

A. DATASET DESCRIPTION

The Food and Drug Administration Adverse Event Reporting System (FAERS) serves as a complete pharmacovigilance database that gathers reports about adverse drug reactions (ADRs) and drug-drug interactions (DDIs) from healthcare professionals together with pharmaceutical manufacturers and consumers. Each report includes complete drug information which includes active substances together with drug name, indications and route of administration and dosage and reported adverse effects and severity classifications. The analysis of medication interactions requires these essential characteristics to study multiple drugs that patients take at the same time.

A major component of the FAERS dataset is the drug name, which identifies the pharmaceutical product involved in the reported adverse event. The drug names used in the FAERS dataset fail to provide complete information about a medication's complete properties. The study incorporates active substances as biologically active chemical compounds that determine therapeutic drug effects to improve prediction accuracy. Active substances play a vital role in understanding drug behavior because they determine both the drug's metabolic processes and its duration in the body and its reactions with other medications.

Another key feature in the dataset is drug indications, which describe the specific diseases or medical conditions a drug is prescribed to treat. Different drugs may share the same indication but have varying mechanisms of action, which can lead to different interaction profiles. Additionally, a single drug may have multiple approved indications, influencing how it interacts with co-administered medications. Our study leverages this information to improve semantic embedding representations, ensuring that drug interactions are not analyzed in isolation but in the broader context of therapeutic use cases.

The adverse event severity classification in FAERS plays a crucial role in assessing the risk associated with drug combinations. The dataset divides drug interactions into serious and non-serious events that include hospitalization or life-threatening reactions and disability and congenital anomalies and death. Machine learning models need these

TABLE 1. Comparative performance of current DDI approaches.

Ref	Objective	Dataset	Applied Algorithm	Performance
[13]	Predicting DDIs using Deep Neural Network.	DrugBank	SMILES (simplified molecular-input line-entry system), a deep neural network model.	The current drugs for treating inflammatory bowel disease (IBD) analyzed, discussing potential drug combinations & providing insights on drug repurposing & development guidelines, achieving an accuracy of 0.932
[14]	Predicting unknown DDIs using Integrated Similarity	DrugBank SIDER KEGG PubChem OFFSIDES	NDD: Neural network-based method for DDI prediction	NDD achieved superior performance in cross-validation with AUPR ranging from 0.830 to 0.947, AUC from 0.954 to 0.994 and F-measure from 0.772 to 0.902.
[15]	Predicting DDI-associated events by Deep Learning	DrugBank	Multimodal Deep Learning framework DDIMDL	With the grouping of sub-structures, targets & enzymes, DDIMDL achieved an accuracy of 0.8852 & an area under the precision-recall curve of 0.9208.
[16]	Investigating DDI via the associations between genes that two drugs target.	DrugBank, KEGG, OSCAR, VA NDF-RT, HPRD, BioGRID, IntAct, & HitPredict.	A ML framework	Two drugs easily interact when targeting common genes, short protein-protein interaction networks, or cross-talking signaling pathways, providing biological insights into potential adverse drug reactions, achieving an accuracy of 0.9479
[17]	Predicting DDIs	DrugBank	Interaction Prediction Graph Attention Network (IPGAT)	The results increase 6.9% in AUROC & at least 8.5% in AUPR for the retrospective experiment.
[18]	Offering a cold start prediction model for both single-type and multiple-type DDIs	drugbank_v5_stanfordlp.npz	CSMDDI, a cold start prediction model for both single-type and multiple-type DDIs	CSMDDI achieves a good performance of DDI prediction in case of both the occurrence prediction and the multi-type reaction prediction in cold start scenario. (Multiple-type DDIs → AUC 0.8658) (Single-type DDIs AUC → 0.8861)
[19]	Providing an AI & ML field-based early detection of ADRs & drug-induced toxicity	List of important databases.	A wide range of methodologies from Data Mining to Deep Learning.	AI and ML in early ADR and toxicity discovery have promising prospects for improving medication safety, reducing medical expenses, and potentially saving lives.
[20]	Predicting adverse drug reaction	DrugBank, SIDER, Decagon	Adaptive Dual Graph Contrastive Learning (ADGCL)	ADGCL is proven to be a promising ADR prediction model, through their overall exploratory outcomes on real-world datasets. F1 0.9667±0.0065
[21]	Offering a deep attributed embedding based multi-task learning model.	FAERS	A Deep Attributed Embedding based Multitask (DAEM) learning model	The DAEM effectiveness is proved when compared with 13 baselines & its variants. DAEM achieves an accuracy of 0.897 ± 0.042
[22]	Predicting DDI using joint deep convolutional transform learning;	DrugBank Database, Stanford's Biosnap dataset.	DeConDFFuse Framework	DeConDFFuse framework performs superior to the benchmarks, achieving an accuracy of 0.907422
[23]	Predicting side effects of a pair of drugs using drug information.	MEDGen	CentSmoothie, a hypergraph neural network (HGNN)	The performance advantages of CentSmoothie were proved in simulations & real datasets.
[24]	Exploring the ChatGPT effectiveness in predicting and explaining common DDIs.	40 DDIs lists from previously published literature.	Descriptive and inferential statistics, Fisher's exact test, unpaired t-test and GraphPad Prism.	ChatGPT, while useful for patients without immediate healthcare access, may provide incomplete guidance on DDIs, requiring further improvement for improved usage and patient information.
[25]	Designing the GNN architecture for DDI prediction without manual intervention	MEDGen	Automatic DDI prediction method named AutoDDI and Reinforcement Learning Search	AutoDDI outperforms two real-world datasets, proving its effectiveness in capturing drug substructure for DDI prediction through visual interpretation.
[26]	Comprising the Drug Chemical Structure Graph-based component and the Drug Knowledge Graph-based component.	DrugBank and KEGG	MSKG-DDI	MSKG-DDI exceeds other models in Binary-class, (an accuracy of 0.9836 for DrugBank & 0.9235 for KEGG); Multi-Class (an accuracy of 0.9614) & multi-label prediction tasks under both transductive & inductive settings & the ablation analysis confirms its functional benefit.
[27]	Providing a more whole understanding of the feature supplement methods used in DDI mining	List of datasets.	Methods based on Deep Learning	The feature supplement methods are compared, & some suggestions are given to approach the current problems & future research directions.

classifications to label drug interactions between high-risk and low-impact drugs.

Our study examined various therapeutic drugs which encompass antibiotics alongside anticoagulants and

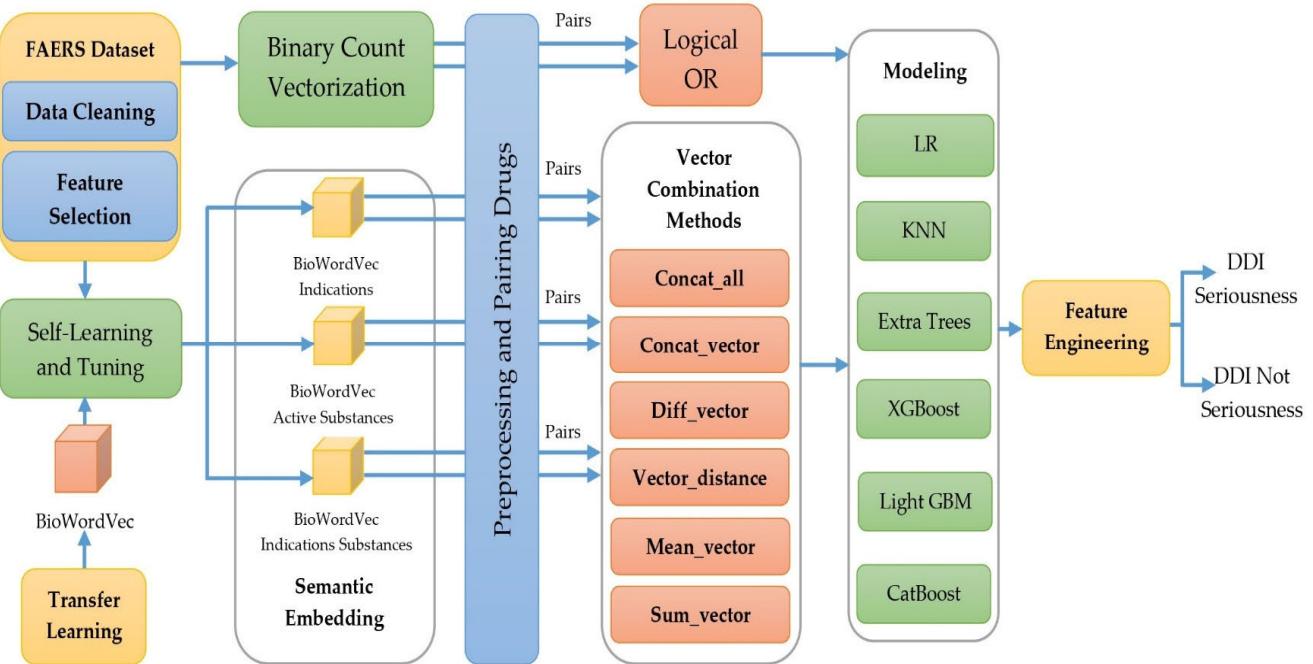


FIGURE 1. Proposed methodology for text vectorization, semantic embedding, and advanced ML.

cardiovascular medicines and antidepressants and antiepileptic drugs and chemotherapy agents that show different pharmacological interaction risks when used with other medications. Warfarin (an anticoagulant) triggers severe bleeding issues with many medications until it becomes deadly while antidepressants such as SSRIs can trigger serotonin syndrome when used with other serotonergic drugs. By analyzing these interactions, our model captures the real-world impact of polypharmacy and helps predict DDI severity with improved accuracy.

B. DATASET PREPARATION

In this paper, the Food and Drug Administration Adverse Event Reporting System (FAERS) dataset [32] is applied. With over 5.9 million recorded events, it stands as a crucial dataset for extensive research in drug safety. This dataset is particularly useful for exploring DDIs and their potential seriousness when two drugs are used concurrently. It offers exhaustive details about each recorded incident, which is instrumental for identifying new safety signals and tracking adverse event trends. The key features of FAERS dataset are described as follows:

- Temporal Scope: The dataset spans from 1989 to 2024, providing a wide temporal window to study long-term trends and historical patterns in drug usage and the associated adverse events.
- Drug Information: Detailed entries including drug names, drug indication, and active substances, critical

for identifying interactions between specific drugs and their linked adverse effects.

- Demographic Data: Includes demographic details such as patient gender, important for understanding how different drugs interact and the varying effects they may have on different populations, thereby supporting tailored therapeutic decisions and safety precautions.

C. DATA CLEANING

To ensure the quality and reliability of our modeling, the initial step involves rigorous data cleaning processes:

- Duplicate Record Removal: A total of 169,507 duplicate records were identified and removed, resulting in a final dataset size of 5,778,106 records. This step was crucial to ensure that only unique records were retained, reducing redundancy and improving the quality of the data used for modeling.
- Handling Missing Values: Records with missing values in critical fields were removed to maintain dataset integrity. The dataset size was reduced from 5,948,483 records to 5,947,613 records after the removal of null values in essential features.

D. FEATURE SELECTION

The selection of relevant features is based on their importance in predicting DDIs:

- Primary Drug Features: medicinal product and Seriousness are selected to identify the drugs involved in the interactions and the severity of the adverse events.

- Extra useful features: drug indication and active substance name are crucial for understanding the context of the drug usage.

E. SECOND PREPROCESSING OF DATA: PAIRING PRODUCTS TOGETHER

In this stage, the dataset is filtered to identify drugs taken together by the same patient at the same time. From the initial set of 5,947,613 entries, this criterion narrows down the data to 60,842 rows, comprising 18,191 entries marked as serious and 42,651 as non-serious. A pair of drugs along with their associated seriousness is then constructed. This step is essential for the next phase of analysis, which involves predicting the seriousness of interactions between these paired drugs. This comprehensive dataset preparation phase sets the foundation for subsequent analyses, enabling us to accurately model and predict the seriousness of drug-drug interactions using advanced computational techniques.

F. TEXT VECTORIZATION USING BINARY COUNT VECTORIZATION

To effectively analyze the textual data within the FAERS database, we employ Binary Count Vectorization. This technique is a variant of the traditional count vectorization method, where instead of counting the frequency of each term, it simply denotes the presence (1) or absence (0) of terms within a document. This method is particularly suited for datasets where the mere presence of a term (e.g., a specific drug name) is more significant than the number of times it appears. Binary Count Vectorization transforms textual data into a binary matrix of token presence:

- Let D be a set of drug reports, and T be the total number of unique terms across all reports.
- Define a vector $V(d)$ for each drug d such that:

$$V(d) = [v_1, v_2, \dots, v_T] \quad (1)$$

where each v_i is defined as:

$$v_i = \begin{cases} 1, & \text{if } i \text{ is present in document representing drug } d \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

- This vectorization approach ensures that each drug is represented as a vector in a high-dimensional space where each dimension corresponds to a unique term from the dataset.

G. APPLICATION IN DRUG INTERACTION ANALYSIS

For analyzing drug interactions, each pair of drugs is represented by combining the vectors of the two individual drugs involved in the interaction. The combination used here is the logical OR operation, which is suited for capturing the presence of any unique term from either drug:

- Let $V(d_1)$ and $V(d_2)$ be the binary vectors representing drugs d_1 and d_2 , respectively.

- The vector representation $V(d_{pair})$ of the drug pair is computed as:

$$V(d_{pair}) = V(d_1) \vee V(d_2) \quad (3)$$

where V denotes the element-wise logical OR operation between the two vectors.

- Each entry in $V(d_{pair})$ thus becomes 1 if the corresponding term is present in either d_1 or d_2 , effectively capturing all terms associated with either drug in the pair.

This binary vector representation of drug pairs is used as input features for ML algorithms. By representing drug interactions in this manner, the model can efficiently learn from the presence of certain drug combinations within the dataset to predict potential DDIs and their seriousness. This approach simplifies the input feature space while preserving crucial information about drug interactions, making it a powerful tool for predictive analytics in pharmacovigilance.

IV. SEMANTIC EMBEDDING USING BIOWORDVEC

The Word2Vec efficiently produces a vector space from a large corpus of text. By using either the Continuous Bag of Words (CBOW) or Skip-Gram models, Word2Vec captures the context of a word in a high-dimensional vector space. In pharmacovigilance, such embedding are crucial for capturing the context in which drugs are mentioned, relating to their indications, interactions, and effects. Semantic embedding plays a critical role in understanding and capturing the nuanced relationships between drug names. BioWordVec, a sophisticated word embedding technique, utilizes pre-trained vectors that incorporate sub-word information from extensive biomedical literature sources such as PubMed and MeSH. This model improves biomedical word embedding with sub-word information and MeSH [30] is particularly adept at encoding semantic similarities and relationships between biomedical terms.

A. EMBEDDING MODEL

Each drug name in the FAERS dataset is mapped to a 200-dimensional vector using BioWordVec based on the following equation:

$$V_{drug} = BioWordVec(drug\ name) \quad (4)$$

where V_{drug} is a 200-dimensional vector representing the semantic embedding of the drug.

Using BioWordVec allows our research to leverage deep semantic learning from one of the most expansive sets of biomedical data available, enhancing our ability to predict and analyze the seriousness of drug-drug interactions.

B. VECTOR COMBINATION METHODS (PAIRS OF DRUGS)

To effectively analyze pairs of drugs for potential interactions and their seriousness, we employ various vector combination methods using vectors derived from WordVec embedding. Each combination method captures different aspects of drug

interactions and provides a distinct view that can be instrumental in modeling with ML algorithms.

- Combining Methods

The concatenation of vectors (`concat_vectors`) is presented in the following equation:

$$V_{concat} = [V_{d1} : V_{d2}] \quad (5)$$

This method simply concatenates the vectors of two drugs, preserving all individual characteristics of each drug. This is particularly useful in models where interactions of specific features between drugs are critical.

- Vector Difference (`diff_vector`)

$$V_{diff} = [V_{d1} - V_{d2}] \quad (6)$$

Applying vector subtraction highlights the differences between the two drug vectors. This method can be beneficial to identify contrasting properties between drugs, which may predict adverse interactions.

- Vector Summation (`sum_vector`)

Vector summation combines the features of both drugs into a single vector, enhancing shared properties. This is useful in models that need to emphasize the combined effect of both drugs.

$$V_{sum} = [V_{d1} + V_{d2}] \quad (7)$$

- Vector Mean (`mean_vector`)

The mean vector calculates the average of each corresponding element from two drug vectors. It provides a balanced view of both drugs, which is effective for capturing the average effect of drugs when combined.

$$V_{mean} = mean[V_{d1}, V_{d2}] \quad (8)$$

- Element-wise Maximum (`max_vector`)

Computes the maximum value for each dimension across the two vectors. This method focuses on the strongest signal from either drug, useful for highlighting dominant drug features that might drive interactions.

$$V_{max} = max[V_{d1}, V_{d2}] \quad (9)$$

- Distance Metrics (`distance`)

Distance methods like Euclidean, Manhattan, and cosine distances measure how far apart the two drug vectors are, providing insights into how dissimilar the drugs are. These metrics are useful in models that need to quantify the degree of difference or similarity between drugs, which can influence interaction effects.

$$V_{euclidean} = \sqrt{\sum_{i=1}^n (V_{d1i} - V_{d2i})^2} \quad (10)$$

$$V_{manhattan} = \sum_{i=1}^n |V_{d1i} - V_{d2i}| \quad (11)$$

$$V_{cosine} = \frac{V_{d1} \cdot V_{d2}}{|V_{d1}| * |V_{d2}|} \quad (12)$$

- Concatenation of All Features (`concat_all`)

This method combines the concatenated vectors with distance metrics, offering a comprehensive feature set that includes both individual and relational information between the drugs.

$$V_{all} = [V_{concat}; V_{euclidean}; V_{manhattan}; V_{cosine}] \quad (13)$$

Each of these vector combination methods provides a unique set of features that may contribute to the next stage of modeling using ML. By experimenting with different combination methods, we can determine which ones capture the most predictive features for the seriousness of drug interactions, thereby enhancing the accuracy and effectiveness of our predictive models.

C. TRANSFER LEARNING ON BIOWORDVEC

While BioWordVec provides a robust foundation for capturing semantic relationships in biomedical contexts, it covers only about 40% of the product names found in the FAERS dataset. Specifically, out of 48,831 unique product names, BioWordVec initially supports only 19,740, leaving 29,001 names uncovered. To address this gap and enhance the model's applicability to our specific dataset, we employ a transfer learning approach, adapting BioWordVec by further training it with drug names, indications, and active substances present in FAERS.

1) TRAINING PROCEDURE

The adaptation uses the skip-gram model from the Word2Vec family, which is designed to predict surrounding words given a current word. This model is particularly suited for enriching embedding with specific domain knowledge:

- Base Model Utilization: We begin with the pre-trained BioWordVec embedding as our base. These embeddings already contain rich biomedical semantics learned from extensive biomedical literature.
- Incorporating FAERS Specific Data: The embedding model is then fine-tuned on the FAERS dataset, focusing on the 29,001 initially uncovered drug names. This fine-tuning process integrates the unique context and usage patterns of these drugs as found in the adverse event reports.

2) DEVELOPMENT OF SPECIALIZED EMBEDDING

To further tailor the embedding to our specific analytical needs, we generate three specialized versions of BioWordVec by integrating different feature sets from the FAERS data:

- BioWordVecIndications: This version is built by combining drug names with their corresponding indications, aiming to capture and encode the relationship between drugs and the conditions they are prescribed for, enriching the semantic understanding of drug use in specific medical contexts.
- BioWordVecActiveSubstance: This model integrates drug names with active substances. The inclusion of active substance information aims to deepen the model's

grasp of the pharmacological aspects of the drugs, which is crucial for understanding potential interactions and effects.

- BioWordVecIndicationSubstance: The most comprehensive of the three, this version combines drug names with both their indications and active substances. By synthesizing all available drug-related data, this embedding model provides the richest semantic representation, designed to capture the full spectrum of drug characteristics and their interactions.

The mathematical model for training these specialized embedding is based on the skip-gram architecture:

- Objective Function:

$$\max \frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log p(w_{t+j}|w_t) \quad (14)$$

where, T is the total number of words in the training corpus, c is the context window size, w_{t+j} are the context words, and w_t is the target word.

- Probability Estimation:

$$p(w_o|w_I) = \frac{\exp(v_w^{IT} v_{wI})}{\sum_{w=1}^W \exp(v_w^{IT} v_{wI})} \quad (15)$$

where, w_o represents the output word, w_I is the input word, and v_w and v_w^l are the ‘input’ and ‘output’ vector representations of word w , with W being the total number of unique words in the vocabulary.

3) SELF-SUPERVISED LEARNING APPROACH

The adaptation and enhancement of BioWordVec are achieved through a self-supervised learning framework. This framework leverages the natural co-occurrence of terms in the dataset to learn more contextualized embedding without the need for explicit labeling, which is often scarce in biomedical datasets:

- Training Dynamics: During training, the skip-gram model adjusts the embedding based on the prediction errors it makes, gradually improving its ability to forecast the correct contextual terms. This iterative refinement helps the embedding to better reflect the specific usage patterns and associations found in the FAERS data.
- Outcome: The result is a set of enriched word embedding that are significantly more aligned with the real-world application of these drugs, thereby enhancing the model’s predictive power and reliability in identifying and assessing drug-drug interactions.

This enhanced transfer learning strategy ensures that our models are not only informed by general biomedical knowledge but are also deeply customized to reflect the specific characteristics and interaction dynamics of the drugs reported in the FAERS dataset.

D. ML MODELS FOR PREDICTING SERIOUSNESS OF DDIs EMBEDDING MODEL

To predict the seriousness of drug interactions effectively, we deploy a range of ML algorithms, each suited to handling the complexity and nuances of the pharmacological data provided by the FAERS dataset. By incorporating traditional and ensemble ML techniques, we aim to leverage the strengths of various approaches to enhance predictive accuracy and reliability. The following ML algorithms are utilized for their respective strengths in classification tasks within the pharmaceutical domain:

- Logistic Regression: A foundational statistical model that estimates probabilities using a logistic function, ideal for binary classification tasks such as predicting the seriousness of DDIs.
- K-Nearest Neighbors (KNN): A non-parametric method that classifies new cases based on a similarity measure (e.g., distance functions), providing robust predictions that adapt well to complex pattern recognition in drug interactions.
- Gaussian Naive Bayes: Based on Bayes’ theorem, this algorithm is particularly effective when assumptions of independence hold, useful for large datasets with many features as in drug interaction analysis.
- XGBoost Classifier: An implementation of gradient boosted decision trees designed for speed and performance, which is highly effective for structured data like ours with both categorical and continuous features.
- LightGBM Classifier: A fast, distributed, high-performance gradient boosting framework based on decision tree algorithms, known for handling large amounts of data efficiently.
- CatBoost Classifier: An algorithm that uses gradient boosting on decision trees, with a specific focus on categorical variables and providing state-of-the-art results for datasets with complex relationships.
- Extra Trees Classifier: An ensemble learning method fundamentally similar to random forests. This algorithm constructs a multitude of decision trees and outputs the class that is the mode of the classes (classification) of the individual trees, offering excellent generalization.

E. FEATURE ENGINEERING

To further enhance model predictions, demographic features such as patient age and sex are integrated with drug interaction vectors. This integration allows for a more nuanced understanding of how demographic factors may influence the seriousness of drug interactions, thus providing tailored predictions.

1) MODEL EVALUATION

All models undergo a rigorous evaluation process using a 5-fold cross-validation approach. This method splits the data into five parts, with each part being used as a test set at different times and the rest serving as a training set. This

ensures the robustness and generalizability of our findings across different subsets of data.

In our evaluation approach, we applied a 5-fold cross-validation strategy to ensure robust assessment of the model. During each fold, the training and testing datasets were kept strictly separate, ensuring that no information from the test set was used during the training phase. Importantly, the fine-tuning of BioWordVec embedding was carried out using the self-supervised learning approach described in Section IV-C3, which utilized data that was not included in the model training process. Specifically, the fine-tuning leveraged additional information from active substance names and drug indications that were distinct from the training set.

This approach ensured that embedding were enriched with relevant domain information without any overlap with the data used for training and testing in each fold. As a result, the evaluation remained unbiased, with the model being tested solely on unseen data throughout the cross-validation process.

2) PERFORMANCE METRICS

The effectiveness and accuracy of each model are assessed using several key performance metrics:

- Accuracy: Measures the proportion of correctly predicted instances to the total instances.
- Precision: Evaluates the accuracy of the positive predictions.
- Recall: Captures the ability of a model to find all the relevant cases (positive class) within a dataset.
- F1 Score: Provides a balance between precision and recall, useful in cases of uneven class distribution.
- ROC AUC: Reflects the model's ability to discriminate between positive and negative classes across different thresholds.

The combination of these diverse algorithms, coupled with robust evaluation techniques, enables a comprehensive analysis of the potential seriousness of DDIs, facilitating better-informed clinical decisions and contributing to overall drug safety.

V. EXPERIMENTAL SETUP

Our study utilizes a structured experimental setup to rigorously evaluate the effectiveness of various ML models in predicting the seriousness of drug-drug interactions (DDIs) using the FAERS dataset. This setup is designed to systematically assess the performance of baseline and advanced models under different conditions.

A. BASELINE MODEL EXPERIMENT

- **Description:** This experiment serves as the baseline for our study. It involves using Count Vectorization to transform drug interaction text data into numerical vectors. This method is tested both with and without the integration of demographic features such as patient age and sex.
- **ML Algorithms:** The transformed data is used to train and test all mentioned ML algorithms, including

Logistic Regression, K-Nearest Neighbors, Gaussian Naive Bayes, XGBoost Classifier, LightGBM Classifier, CatBoost Classifier, and Extra Trees Classifier.

- **Evaluation Metrics:** Performance is measured using accuracy, precision, recall, F1 score, and ROC AUC to establish a comprehensive understanding of each model's capabilities.

B. ADVANCED MODEL EXPERIMENTS

The advanced model experiments leverage BioWordVec and its specialized variants to explore the impact of semantic embedding on the predictive accuracy:

- **Experiment 2: BioWordVec Embedding**
 - **Description:** This experiment utilizes the general BioWordVec embedding to encode drug names. Different vector combination methods are tested, such as vector concatenation, sum, and element-wise maximum, with and without demographic features.
 - **Objective:** To determine the effectiveness of BioWordVec embedding in capturing the semantic nuances of drug names and their impact on predicting DDIs.
- **Experiment 3: BioWordVecIndications Embedding**
 - **Description:** This setup employs embedding that combine drug names with their indications, aiming to capture contextual information related to the conditions treated by the drugs.
 - **Objective:** To assess how the inclusion of drug indications in the embedding influences the model's ability to predict serious DDIs.
- **Experiment 4: BioWordVecActiveSubstance Embedding**
 - **Description:** In this experiment, embedding that integrate drug names with their active substances are used. The focus is on the pharmacological aspects of the drugs.
 - **Objective:** To evaluate whether the inclusion of active substance information enhances the prediction of DDIs' seriousness.
- **Experiment 5: BioWordVecIndicationSubstance Embedding**
 - **Description:** The most comprehensive experiment uses embedding that combine drug names with both their indications and active substances.
 - **Objective:** To explore the combined effect of including both clinical and pharmacological contexts in the embedding on DDI seriousness prediction.

C. METHODOLOGY FOR ALL EXPERIMENTS

Data Preparation: Each dataset variant is prepared with appropriate vectorization or embedding techniques.

- **Model Training:** Models are trained using a 5-fold cross-validation method to ensure robustness and generalizability.

- Comparative Analysis: Results from baseline and advanced model experiments are compared to identify which methods provide the most significant improvements in predicting the seriousness of DDIs.
- Impact Assessment: The practical implications of employing these models in clinical settings are discussed, considering the potential to enhance patient safety through better prediction of adverse drug interactions.

This experimental setup not only tests the hypothesis that advanced semantic embedding can improve DDI prediction but also establishes a benchmark for future research in pharmacovigilance analytics.

VI. CONDUCTED RESULTS

Our study utilizes a structured experimental setup to rigorously evaluate the effectiveness of various ML models in predicting the seriousness of drug-drug interactions (DDIs) using the FAERS dataset. This setup is designed to systematically assess the performance of baseline and advanced models under different conditions.

In this study, the models' performance was evaluated by aggregating the predictions from all the portions of the test set of the 5-fold cross-validation. In particular, instead of reporting the mean and standard deviation of the metrics for each fold, we assessed the overall performance based on the predictions from all folds, which is a single set of values that would adequately describe the given dataset. This approach gave a more general assessment of the model's performance over the entire data set since it avoided the variation that the evaluation of the different folds in isolation would occasion. However, we understand that reporting fold-wise metrics might help in understanding the performance fluctuations across different folds of the data, and we propose to include these results in future experiments for a better understanding of the stability of the model.

A. EXPERIMENT 1 - COUNT VECTORIZATION WITH AND WITHOUT DEMOGRAPHIC DATA

In Experiment 1, we evaluated the impact of including demographic features (age and sex) in conjunction with Count Vectorization on the prediction of the seriousness of DDIs. The models were assessed based on their F1 scores, using a macro average to account for class imbalance. Here, we present the comparative results of this experiment. As presented in Table 2, a summary of F1 scores achieved by different ML algorithms is explained, both with and without demographic data integration.

Based on the conducted results in Table 1, we can identify the following:

- General Observations: The inclusion of demographic data alongside the high-dimensional count vectorized features (8,348 features) does not significantly alter the model performance. The slight variations in F1 scores when demographic features are added suggest that the demographic data's impact is overshadowed by the more

TABLE 2. F1-score for count vectorization with and without demographic data.

Model	F1 Score without Demographic	F1 Score with Demographic
K-Nearest Neighbors	65.1%	63.7%
LightGBM Classifier	67.3%	67.3%
Extra Trees Classifier	67.4%	67.3%
XGBoost Classifier	67.5%	66.9%
Logistic Regression	67.7%	66.3%
CatBoost Classifier	68.3%	68.4%

dominant drug-related features in the count vectorization model.

- Best Performing Model: The CatBoost Classifier stands out as the most effective model, showing a slight improvement when demographic data is included. It demonstrates robustness and adaptability to both feature sets, indicating its efficacy in handling large feature spaces with complex interactions.
- Stability across Configurations: The LightGBM Classifier and Extra Trees Classifier exhibited consistent performance regardless of the inclusion of demographic data, suggesting that these models are inherently stable and less influenced by minor feature additions.

To identify the interpretation of demographic impact on the results, we can conclude the following:

- Limited Influence of Demographic Features: The modest impact of demographic features on the model performance can be attributed to the overwhelming presence of the 8,348 count vectorization features. In such a high-dimensional feature space, the addition of just four demographic features likely does not provide substantial new information to significantly shift model outcomes.
- Model Dependency on Feature Sensitivity: The differential impact on model performance with the addition of demographic data also highlights how various algorithms manage feature sensitivity and complexity. Models like CatBoost and LightGBM, which are equipped to handle high-dimensional data and complex interactions effectively, are less likely to be significantly influenced by the addition of relatively fewer demographic features.

Experiment 1 highlights the subtle yet significant role of demographic features within a complex, high-dimensional feature space designed to predict the seriousness of drug-drug interactions (DDIs).

While including demographic data typically does not result in substantial performance enhancements, specific sophisticated models can leverage these features to achieve slight improvements.

Notably, this experiment's peak performance was achieved by the CatBoost Classifier, which reached an F1 score of 68.4% without the integration of demographic data. This finding is pivotal for refining feature engineering and model training approaches in future experiments, particularly in

high feature dimensionality scenarios. Such insights are instrumental in enhancing the accuracy and efficacy of predictive models in pharmacovigilance.

B. EXPERIMENT 2 RESULTS: BIOWORDVEC EMBEDDING

Experiment 2 focuses on assessing the impact of using BioWordVec embedding on the prediction of the seriousness of drug-drug interactions (DDIs) without and with the integration of demographic features. Various vector combination methods were explored to determine how different approaches to merging drug information impact the effectiveness of our ML models. The results provide insights into how semantic embedding influence model performance across different configurations.

While BioWordVec addressed only 40% of the drug names in the FAERS dataset, it was selected for this research because of its adaptive domain. BioWordVec is trained from the biomedical literature which gives essential semantic information pertinent to biomedical and pharmacological domains.

Also, recent embedding models like the BERT based models or even the biomedical embedding based on transformers, may need much more computational resources for training and tuning. Preliminary tests showed that BioWordVec had reasonable accuracy and is fast enough to be used for our particular task. We have also employed transfer learning to allow BioWordVec to capture more of the FAERS-specific drug names, leading to an improvement in coverage to over 80% and general performance of the model.

1) THE BIOWORDVEC EMBEDDING RESULTS WITHOUT DEMOGRAPHIC FEATURES

As presented in Table 3, the results summarize the performance of various ML algorithms using different vector combination methods without incorporating demographic data.

The highest performances were achieved by the CatBoost Classifier using concat_all (72.27%), max_vector (72.23%), and concat_vectors (72.21%). These results underscore the effectiveness of the CatBoost algorithm in handling complex feature interactions without demographic data.

2) THE BIOWORDVEC EMBEDDING RESULTS WITH DEMOGRAPHIC FEATURES

When demographic data is included, the following results presented in Table 4 were observed across the same vector combination methods.

The top results were also by the CatBoost Classifier, showing slightly improved scores: concat_all (72.88%), max_vector (72.75%), and concat_vectors (72.74%). This indicates a marginal improvement with the addition of demographic features, suggesting that these features may provide some contextual benefit that slightly enhances the predictive accuracy of the model.

Experiment 2 demonstrates a significant enhancement in model performance with the introduction of BioWordVec embedding, especially when compared to Experiment 1,

which utilized count vectorization. The CatBoost Classifier consistently excelled, achieving superior outcomes under both conditions with complex vector combination methods such as concat_all, concat_vectors, and max_vector. This improvement is particularly striking, with the best performance reaching 72.88%, surpassing the top F1 score of 68.4% from Experiment 1. The inclusion of demographic data provided only a marginal boost in performance, emphasizing the nuanced role these features play when integrated with sophisticated embedding techniques. While demographic features alone do not transform model outcomes, their subtle contributions significantly enhance the models' ability to discern the seriousness of drug-drug interactions with greater precision when used in conjunction with BioWordVec embedding. This experiment underlines the effectiveness of semantic embedding in improving the accuracy of predictions in pharmacovigilance. It advocates for integrating advanced ML techniques with enriched feature sets, including semantic embedding and contextual and demographic data, to optimize the predictive modelling of DDIs.

These advancements are pivotal for future research and practical applications, highlighting the potential of merging deep learning techniques with traditional data to enhance outcome predictions in medical informatics.

C. EXPERIMENT 3: BIOWORDVECINDICATIONS EMBEDDING

Experiment 3 investigates the use of BioWordVecIndications embedding, which combines drug names with their indications, to predict the seriousness of drug-drug interactions (DDIs). The effectiveness of various ML models was tested using different vector combination methods, both with and without the integration of demographic features, to determine how the enriched embedding impact model performance.

1) THE BIOWORDVECINDICATIONS EMBEDDING RESULTS WITHOUT DEMOGRAPHIC FEATURES

As presented in Table 5, the performance of various ML algorithms is summarized using different vector combination methods without incorporating demographic data. The highest performances were achieved by the CatBoost Classifier using concat_all (72.51%), concat_vectors (72.46%), and max_vector (72.30%). These results underscore the effectiveness of the CatBoost algorithm in handling complex feature interactions without demographic data.

2) THE BIOWORDVECINDICATIONS EMBEDDING RESULTS WITH DEMOGRAPHIC FEATURES

When demographic data is included, the results of Table 6 were observed across the same vector combination methods.

When demographic data was included, the top results were also by the CatBoost Classifier, showing slightly improved scores: concat_vectors (73.11%), concat_all (73.04%), and max_vector (72.85%). This indicates a marginal improvement with the addition of demographic features, suggesting

TABLE 3. Biowordvec embedding without demographic features.

Vector Combination Method	Gaussian Naive Bayes	Logistic Regression	K-Nearest Neighbors	Extra Trees Classifier	XGBoost Classifier	LightGBM Classifier	CatBoost Classifier
concat_all	64.25%	67.45%	68.48%	69.70%	71.29%	71.79%	72.27%
concat_vectors	64.24%	67.54%	69.03%	69.64%	71.36%	71.73%	72.21%
diff_vector	59.91%	53.47%	68.56%	67.02%	68.76%	68.81%	70.20%
distance	40.61%	53.10%	63.40%	66.31%	62.41%	62.43%	62.31%
max_vector	63.58%	66.07%	68.84%	69.20%	71.20%	71.57%	72.23%
mean_vector	64.16%	67.71%	68.97%	68.24%	70.19%	70.12%	71.21%
sum_vector	64.16%	67.72%	68.97%	68.24%	70.19%	70.12%	71.21%

TABLE 4. Biowordvecindications embedding with demographic features.

Vector Combination Method	Gaussian Naive Bayes	Logistic Regression	K-Nearest Neighbors	Extra Trees Classifier	XGBoost Classifier	LightGBM Classifier	CatBoost Classifier
concat_all	64.18%	65.88%	65.96%	68.55%	71.93%	72.34%	72.88%
concat_vectors	64.26%	66.40%	67.17%	68.50%	72.13%	72.31%	72.74%
diff_vector	60.59%	53.76%	65.86%	65.74%	69.18%	69.38%	70.92%
distance	54.68%	53.42%	58.53%	63.92%	63.67%	63.29%	64.12%
max_vector	63.81%	65.02%	66.97%	67.78%	71.85%	72.13%	72.75%
mean_vector	63.61%	65.82%	67.17%	66.98%	70.86%	71.01%	71.96%
sum_vector	63.90%	66.24%	67.50%	66.98%	70.86%	71.01%	71.96%

that these features may provide some contextual benefit that slightly enhances the predictive accuracy of the model.

Experiment 3 demonstrates a notable enhancement in model performance by introducing BioWordVecIndications embedding, which combines drug names with their indications. This approach contrasts sharply with results from previous experiments that used less sophisticated embedding methods. The CatBoost Classifier consistently excelled, achieving superior outcomes under both conditions with complex vector combination methods such as concat_all, concat_vectors, and max_vector. This improvement is particularly significant, with a top performance reaching 73.11%, surpassing the highest F1-score of 72.88% from Experiment 2.

This underscores the substantial impact of utilizing enriched word embedding over traditional methods, illustrating the advantages of deeper semantic integration in model accuracy. Including demographic data provided only a marginal boost in performance, emphasizing the nuanced role these features play when integrated with advanced embedding techniques. While demographic features alone do not transform model outcomes, their subtle contributions significantly enhance the models' ability to discern the seriousness of drug-drug interactions more precisely when used alongside BioWordVecIndications embedding.

The results demonstrate the utility of semantic embedding for improving prediction accuracy in pharmacovigilance. It proposes integrating sophisticated ML techniques with enriched feature sets such as semantic embedding and contextual and demographic data to maximize the predictive modeling of DDIs. The development of these advancements will play a pivotal role in future research and practical application in medical informatics, as it suggests the potential of combining deep learning techniques with more traditional data to improve outcome prediction.

D. EXPERIMENT 4: BIOWORDVEACTIVE SUBSTANCE EMBEDDING

Experiment 4 evaluates the effectiveness of BioWordVecActiveSubstance embedding, which combines drug names with active substances, on the prediction of the seriousness of drug-drug interactions (DDIs). This experiment tested various vector combination methods with ML models, assessing performance both with and without the inclusion of demographic data.

1) THE BIOWORDVEACTIVE SUBSTANCE EMBEDDING RESULTS WITHOUT DEMOGRAPHIC FEATURES

Table 7 summarizes the performance of different ML models using various vector combination methods, without incorporating demographic data.

The CatBoost Classifier demonstrated the highest performance across multiple vector combination methods, notably achieving 72.42% with concat_all, 72.28% with concat_vectors, and 71.93% using LightGBM Classifier with max_vector.

2) THE BIOWORDVEACTIVE SUBSTANCE EMBEDDING RESULTS WITH DEMOGRAPHIC FEATURES

The inclusion of demographic data provided slight improvements in some models, particularly in the CatBoost Classifier, which again showed the highest scores with concat_all (73.21%), concat_vectors (73.13%), and max_vector (72.61%) as explained in Table 8.

Experiment 4 demonstrated a significant advancement in predictive accuracy through the use of BioWordVecActiveSubstance embedding, which integrates active substance data into drug names. This new embedding approach led to the CatBoost Classifier reaching a peak performance of 73.21% with demographic data using the concat_all

TABLE 5. Biowordvecindications embedding without demographic features.

Vector Combination Method	Gaussian Naive Bayes	Logistic Regression	K-Nearest Neighbors	Extra Trees Classifier	XGBoost Classifier	LightGBM Classifier	CatBoost Classifier
concat_all	60.52%	69.01%	67.29%	69.52%	72.01%	72.00%	72.51%
concat_vectors	60.54%	69.32%	68.25%	69.66%	71.92%	71.94%	72.46%
diff_vector	59.55%	54.01%	67.78%	67.15%	68.69%	68.47%	69.97%
distance	41.18%	53.80%	62.85%	66.74%	62.52%	61.96%	61.78%
max_vector	61.04%	68.25%	68.59%	69.65%	71.63%	71.24%	72.30%
mean_vector	59.84%	69.28%	68.47%	68.90%	70.77%	70.30%	71.38%
sum_vector	59.84%	69.34%	68.47%	68.90%	70.77%	70.30%	71.38%

TABLE 6. Biowordvecindications embedding with demographic features.

Vector Combination Method	Gaussian Naive Bayes	Logistic Regression	K-Nearest Neighbors	Extra Trees Classifier	XGBoost Classifier	LightGBM Classifier	CatBoost Classifier
concat_all	60.53%	67.25%	66.10%	68.23%	72.42%	72.42%	73.04%
concat_vectors	60.58%	67.00%	67.17%	68.16%	72.33%	72.40%	73.11%
diff_vector	59.39%	54.11%	65.95%	65.60%	69.39%	69.21%	70.67%
distance	57.38%	53.73%	59.22%	64.30%	63.69%	63.09%	63.98%
max_vector	60.78%	67.44%	67.07%	68.27%	72.08%	72.01%	72.85%
mean_vector	59.77%	67.68%	67.18%	67.44%	71.15%	70.84%	72.08%
sum_vector	59.76%	67.94%	67.53%	67.45%	71.15%	70.84%	72.08%

TABLE 7. Biowordvecactivesubstance embedding without demographic features.

Vector Combination Method	Gaussian Naive Bayes	Logistic Regression	K-Nearest Neighbors	Extra Trees Classifier	XGBoost Classifier	LightGBM Classifier	CatBoost Classifier
concat_all	59.06%	67.69%	66.68%	69.26%	71.67%	71.73%	72.42%
concat_vectors	59.09%	67.89%	67.80%	69.27%	71.91%	71.93%	72.28%
diff_vector	56.36%	54.20%	67.33%	67.00%	68.27%	67.97%	69.35%
distance	41.19%	53.79%	63.06%	66.82%	62.55%	62.71%	62.77%
max_vector	59.48%	66.74%	68.19%	69.30%	71.10%	70.91%	71.83%
mean_vector	58.62%	67.82%	68.11%	68.56%	69.77%	69.43%	70.25%
sum_vector	58.62%	67.79%	68.11%	68.53%	69.77%	69.43%	70.25%

TABLE 8. Biowordvecactivesubstance embedding with demographic features.

Vector Combination Method	Gaussian Naive Bayes	Logistic Regression	K-Nearest Neighbors	Extra Trees Classifier	XGBoost Classifier	LightGBM Classifier	CatBoost Classifier
concat_all	58.97%	65.94%	65.50%	67.85%	72.20%	72.39%	73.21%
concat_vectors	59.04%	65.48%	66.61%	67.75%	72.35%	72.41%	73.13%
diff_vector	55.40%	53.04%	65.53%	65.50%	68.75%	68.54%	69.98%
distance	56.70%	54.74%	59.30%	63.87%	63.48%	62.77%	63.70%
max_vector	59.25%	65.66%	66.14%	67.91%	72.01%	71.71%	72.61%
mean_vector	58.53%	65.91%	65.96%	67.06%	70.41%	70.13%	70.92%
sum_vector	58.56%	66.25%	66.83%	67.04%	70.41%	70.13%	70.92%

vector combination method. This marks an improvement from Experiment 3, where BioWordVecIndications embedding—combining drug names with their indications—yielded a slightly lower top accuracy of 73.04% for the same model and method. This enhancement underscores the value of including active substance information, which appears to deepen the model's understanding and improve its ability to accurately predict the seriousness of drug-drug interactions.

E. EXPERIMENT 5: BIOWORDVECINDICATIONSUBSTANCE EMBEDDING

Experiment 5 evaluates the effectiveness of BioWordVecIndicationSubstance embedding, which integrates both drug

indications and active substances into drug names, on the prediction of the seriousness of DDIs. This experiment tested various vector combination methods with ML models, assessing performance both with and without the inclusion of demographic data.

1) THE BIOWORDVECINDICATIONSUBSTANCE EMBEDDING RESULTS WITHOUT DEMOGRAPHIC FEATURES

Table 9 highlights the performance of different ML models using various vector combination methods, without incorporating demographic data.

The CatBoost Classifier demonstrated the highest performance across multiple vector combination methods,

notably achieving 72.66% with concat_all, 72.56% with concat_vectors, and 72.15% using LightGBM Classifier with max_vector.

2) THE BIOWORDVECINDICATIONSUBSTANCE EMBEDDING RESULTS WITH DEMOGRAPHIC FEATURES

The inclusion of demographic data provided slight improvements in some models, particularly in the CatBoost Classifier, which again showed the highest scores with concat_vectors (73.32%), concat_all (73.30%), and max_vector (72.88%) as presented in Table 10.

Experiment 5 demonstrated a significant advancement in predictive accuracy through the use of BioWordVecIndicationSubstance embedding, integrating both indications and active substances into drug names. This comprehensive embedding approach led to the CatBoost Classifier reaching a peak performance of 73.32% with demographic data using the concat_vectors vector combination method. This marks an improvement from Experiment 4, where BioWordVecActiveSubstance embedding—combining only active substances—yielded a slightly lower top accuracy of 73.21% for the same model and method.

F. FACTORS CONTRIBUTING TO OPTIMAL PREDICTIVE PERFORMANCE

To clarify the reasons behind achieving the best F1 score of 73.32% and ROC AUC score of 84%, five issues must be explained:

1) FEATURE REPRESENTATION AND SEMANTIC EMBEDDING

One of the key contributions of our study is the development and refinement of BioWordVecIndicationSubstance embedding, which integrates drug names, their indications, and active substances.

This enriched semantic representation enables the model to capture deeper pharmacological relationships between drugs, leading to more accurate classification of drug-drug interaction (DDI) severity.

- The baseline count vectorization approach lacked the ability to capture complex relationships, achieving a lower F1 score (68.4%).
- By incorporating semantic embedding (BioWordVec), the performance improved significantly (F1 score increased to 72.88%).
- The best performance was achieved with BioWordVecIndicationSubstance, where drug names were embedded alongside their medical indications and active substances.

This model provided richer contextual information, leading to the highest F1 score of 73.32%.

2) VECTOR COMBINATION METHODOLOGY

We systematically evaluated various vector combination methods to determine the most effective way to represent drug pairs:

- Concat_vectors was the most effective technique, achieving the highest F1 score. This method preserves individual drug characteristics while also allowing the model to learn their joint interactions.
- Other methods, such as max_vector and mean_vector, showed slightly lower performance, indicating that preserving complete feature sets was more beneficial.

3) MODEL SELECTION AND TRAINING STRATEGY

The CatBoost Classifier demonstrated superior performance compared to other machine learning models due to its ability to:

- High efficiency in handling high-dimensional sparse data.
- Capturing complex nonlinear interactions between drug embedding.
- Optimize feature importance, making it well-suited for the pharmacovigilance dataset.

4) ROC AUC SCORE OF 84% WITH STRONG DISCRIMINATIVE POWER

The ROC AUC score of 84% reflects the model's ability to distinguish between serious and non-serious drug interactions effectively:

- The model effectively identified positive cases (serious DDIs), as indicated by a true positive rate of 71.81%.
- The high specificity (78.17%) ensured a reduced number of false positives.

The area under the curve (AUC) illustrates that the classifier maintains a strong balance between sensitivity and specificity, making it a reliable predictive tool for pharmacovigilance.

5) INFLUENCE OF TRANSFER LEARNING

A crucial improvement was the transfer learning enhancement applied to BioWordVec, which allowed the embedding to cover over 80% of the FAERS dataset drug names (compared to only 40% before fine-tuning). This fine-tuning process provided:

- Better contextual understanding of FAERS-specific drugs.
- Higher coverage of drug entities, reducing sparsity in embedding.
- Improved generalization, making the model more effective for real-world pharmacovigilance applications.

G. DETAILED RESULTS OF THE BEST MODEL: CATBOOST CLASSIFIER

The CatBoost Classifier, employing the concat_vectors method with the inclusion of demographic data, stands out as the best-performing model in our series of experiments. This model demonstrated exceptional effectiveness in predicting the seriousness of drug-drug interactions across various metrics.

TABLE 9. Biowordvecindicationsubstance embedding without demographic features.

Vector Combination Method	Gaussian Naive Bayes	Logistic Regression	K-Nearest Neighbors	Extra Trees Classifier	XGBoost Classifier	LightGBM Classifier	CatBoost Classifier
concat_all	63.63%	68.77%	67.47%	69.51%	72.06%	72.13%	72.66%
concat_vectors	63.57%	69.13%	68.25%	69.54%	72.15%	72.05%	72.56%
diff_vector	60.41%	54.22%	67.85%	67.22%	68.77%	68.45%	69.98%
distance	42.97%	53.56%	63.65%	66.90%	62.06%	62.38%	62.17%
max_vector	63.88%	67.59%	68.90%	69.57%	71.50%	71.27%	72.08%
mean_vector	63.12%	68.96%	68.76%	68.88%	70.62%	70.48%	71.45%
sum_vector	63.12%	69.00%	68.76%	68.89%	70.62%	70.48%	71.45%

TABLE 10. Biowordvecindicationsubstance embedding with demographic features.

Vector Combination Method	Gaussian Naive Bayes	Logistic Regression	K-Nearest Neighbors	Extra Trees Classifier	XGBoost Classifier	LightGBM Classifier	CatBoost Classifier
concat_all	63.60%	67.30%	66.78%	68.15%	72.60%	72.68%	73.30%
concat_vectors	63.52%	67.89%	67.54%	68.24%	72.71%	72.55%	73.32%
diff_vector	60.64%	53.95%	66.70%	65.69%	69.20%	69.15%	70.53%
distance	57.32%	54.73%	59.14%	63.97%	63.50%	62.85%	63.70%
max_vector	63.88%	66.63%	67.50%	68.23%	72.12%	72.20%	72.88%
mean_vector	63.06%	67.78%	67.22%	67.45%	71.24%	71.18%	72.09%
sum_vector	63.09%	67.59%	67.71%	67.42%	71.24%	71.18%	72.09%

1) KEY PERFORMANCE METRICS

The overall key performance metrics of the CatBoost Classifier are explored in Table 11.

These metrics signify a well-balanced model that excels not only in accurately identifying a high percentage of true interactions but also maintains commendable precision and recall, effectively minimizing both false positives and false negatives.

TABLE 11. Performance of catboost classifier.

Predicted Actual	Predicted Negative	Predicted Positive
Actual Negative	78.17%	21.83%

2) CONFUSION MATRIX ANALYSIS

Table 12 explains the confusion matrix for the CatBoost Classifier as follows:

- True Negative Rate (Specificity): The model successfully identified 78.17% of the non-interactions, showcasing its capability to recognize true negative cases accurately.
- True Positive Rate (Sensitivity or Recall): 71.81% of the actual interactions were correctly predicted, highlighting the model's proficiency in detecting true positives.
- False Positive Rate: 21.83% of the non-interactions were incorrectly predicted as interactions.
- False Negative Rate: 28.19% of the actual interactions were missed, pointing out potential improvement areas where the model could capture more subtle or complex interactions more effectively.

As presented in Fig. 2, the ROC AUC Score of 0.84 corroborates the model's strong capability to distinguish between

TABLE 12. Actual and predicted results for catboost classifier.

Perf Classifier	Accuracy	F1 Score	Precision	Recall	ROC AUC Score
CatBoost					
Classifier	76.27%	73.32%	72.55%	74.99%	0.84

classes effectively. A score closer to 1 indicates excellent performance, with the CatBoost Classifier demonstrating robust discriminative ability.

This detailed analysis accentuates the model's strengths in achieving high accuracy and reliability in its predictions, making it an invaluable asset in pharmacovigilance. The results advocate for the continued deployment and further refinement of the CatBoost Classifier, mainly focusing on enhancing areas where reductions in false negatives can improve the model's sensitivity without significantly impacting its precision. Fig. 2 shows the ROC curve of the CatBoost model for classifying serious and non-serious DDI interactions. The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1-specificity) at each classification threshold, showing how well the model classifies between the two categories. The CatBoost model has an AUC (Area under the Curve) of 84%, which is better than random guessing (AUC = 50%) and outperforms many baseline models. The steep rise of the curve towards the top left corner shows that the model has a very high true positive rate while still keeping false positives at a low level. In pharmacovigilance, this is important because missing serious interactions (false negatives) can be as important as finding spurious alerts (false positives). Furthermore, the optimal threshold can be found from the curve to determine the right balance between

sensitivity and specificity, making the model accurate and practical for use in real-world applications.

The steep rise at the beginning of the curve indicates that the model captures a high proportion of true positives early while keeping the false positive rate low, which is crucial in pharmacovigilance applications to ensure that severe interactions are not overlooked. This confirms that CatBoost effectively distinguishes high-risk drug interactions with a strong balance between recall and precision.

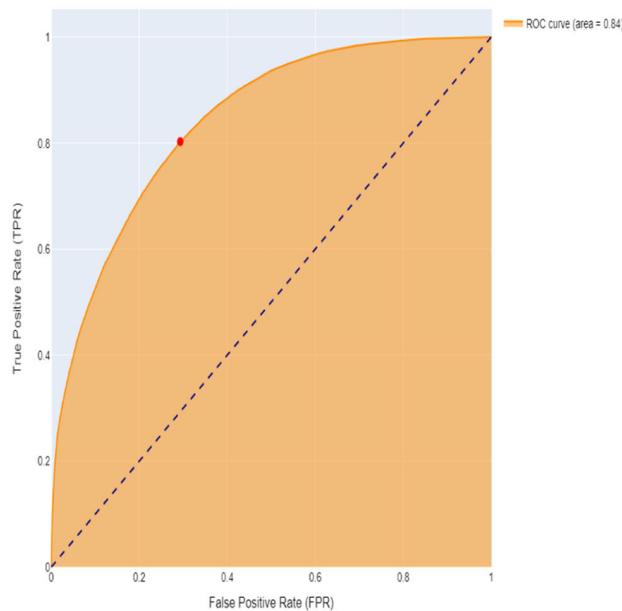


FIGURE 2. ROC curve for catboost classifier.

H. EFFECTIVENESS OF TRANSFER LEARNING FOR BIOWORDVEC

In order to assess the effectiveness of the transfer learning process, we conducted a comparison of the predictive accuracy of the original BioWordVec model with the model retrained on FAERS data. When the demographic features were included, the performance of the un-adapted BioWordVec embedding was at 72.88% F1 score. This performance was already promising but demonstrated its ability to extract specific details of drug-drug interactions from FAERS.

When we performed the transfer learning to fine-tune BioWordVec using the data obtained from FAERS that contains both indication and active substance information, we experienced an increase in accuracy of the model. In particular, the F1 score was enhanced to 73.32% when using the modified BioWordVecIndicationSubstance embedding. This improvement underscore the advantage of including drug-specific contextual information into the embedding model, which enables a more detailed mapping of drug relationships and improved prediction of the seriousness of drug-drug interactions.

The comparison shows that the proposed approach of using transfer learning to adapt BioWordVec for FAERS has produced a slight but significant enhancement in the performance of predicting drug interactions based on the semantic analysis of the data.

VII. DISCUSSION

In addressing the research questions introduced at the beginning of this study, significant advancements were made in predicting the seriousness of drug-drug interactions (DDIs) using advanced data analytics. The findings drawn from various ML models and embedding techniques are particularly notable when analyzed through the lens of the attached comprehensive performance summary.

Question 1: Utilization of Text Vectorization and ML

The study examined the combination of text vectorization and advanced ML models for predicting the seriousness of drug-drug interactions (DDIs). Experiment 1 focused on Count Vectorization with and without demographic data (age and sex). The CatBoost Classifier, when demographic data was not included, achieved the highest F1 score of 68.3%, demonstrating its effectiveness. This underlines the robust capability of ML models, including CatBoost, to predict serious outcomes when multiple drugs are taken concurrently. This finding confirms the effectiveness of computational techniques in a pharmacovigilance context.

Question 2: Role of Semantic Embedding in Predictive Accuracy

Experiment 2 has demonstrated a notable advancement in predictive accuracy for assessing the seriousness of drug interactions, especially when comparing results with those from Experiment 1, which utilized count vectorization. The CatBoost Classifier, employing the concat_all vector combination method with BioWordVec embedding and including demographic data, achieved the highest F1 score of 72.88%. This marks a significant improvement over the top performance from Experiment 1, where the best F1 score was 68.4% using CatBoost without demographic data.

This improvement underscores the substantial impact of semantic embedding in enhancing model performance. Integrating sophisticated ML techniques with enriched semantic embedding and demographic data makes the predictive modeling of drug-drug interactions more accurate and insightful. The resounding success of the CatBoost Classifier with concat_all in Experiment 2, achieving the highest F1 score of 72.88%, highlights the efficacy of this approach and instills confidence in its potential. This benchmark performance advocates for continued exploration of advanced embedding techniques to optimize outcomes in pharmacovigilance.

Question 3: Application of Transfer Learning

Transfer learning has proven highly effective in adapting pre-trained embedding to enhance the predictive accuracy of drug-drug interactions (DDIs), considering specific patient demographics and drug characteristics. The adjustments to the default Word2Vec, which expanded its coverage to include drugs not initially supported by BioWordVec, marked

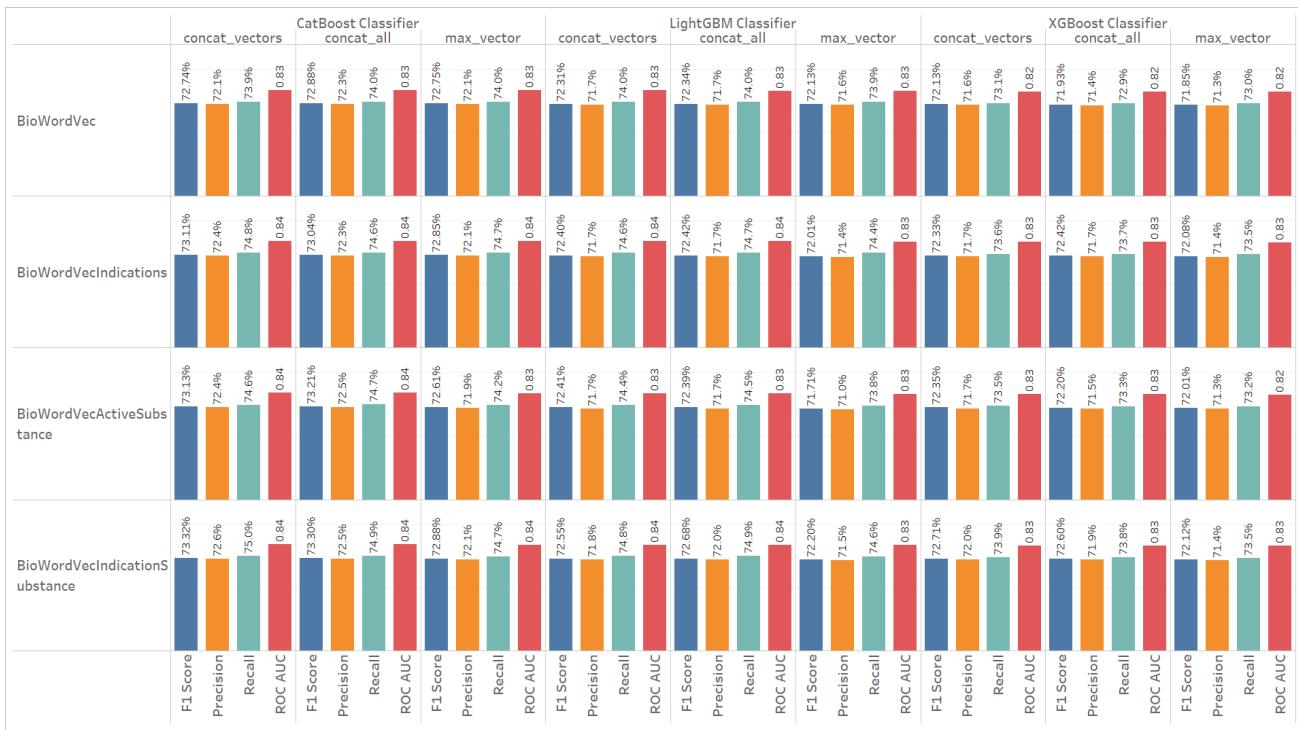


FIGURE 3. Performance metrics across different embedding methods and ML models.

a significant improvement in the model's applicability and performance. Initially, BioWordVec covered only 19,740 drug names, leaving out 29,001 names. The enhancement involved tuning this model to better represent the vast array of drugs in the FAERS dataset. In the experiments:

- Experiment 3 showed remarkable results with the newly enhanced embedding (BioWordVecIndications). The CatBoost Classifier achieved an F1 score of 73.11% using the concat_vectors method, which was the highest in this setting.
- Experiment 4 pushed the boundaries further by integrating active substances into the embedding (BioWordVecActiveSubstance). The CatBoost Classifier peaked with an F1 score of 73.21% using the concat_all method.
- Experiment 5 integrated drug indications and active substances (BioWordVecIndicationSubstance), culminating in the highest performance, with the CatBoost Classifier reaching an F1 score of 73.32% using the concat_vectors method.

These enhancements reflect a substantial leap from BioWordVec's initial capabilities, driven by the application of transfer learning. The tuning of Word2Vec not only broadened the scope of drug names that could be analyzed but also significantly enriched the model's predictive power by capturing more detailed and relevant features from the FAERS dataset. This advanced application of transfer learning improved the adaptability of pre-trained models and their accuracy in predicting serious outcomes of DDIs, providing

a robust tool for enhancing pharmacovigilance and patient safety.

Fig. 3 presents a detailed visual representation of the performance metrics across different embedding methods and ML models.

It serves as a testament to the rigorous analytical approach adopted in this study and underscores several key insights:

- **Top Models:** The CatBoost Classifier has demonstrated superior performance consistently across various embedding techniques and vector combination methods. This model excels in handling complex interactions and intricate data patterns, which is evident from its consistently high scores across F1, precision, and ROC AUC metrics.
- **Best Embedding Techniques:** Among the embedding techniques, BioWordVecIndicationSubstance emerges as the most effective, particularly when paired with the CatBoost Classifier. This embedding method, which integrates both drug indications and active substances, has shown the highest scores in most metrics, highlighting its capacity to capture the multifaceted nature of drug interactions more comprehensively.
- **Performance Metrics Insight:** The precision and ROC AUC metrics are particularly notable with the CatBoost Classifier, emphasizing its robustness in classification accuracy and its ability to distinguish between the classes effectively. This is critical in clinical settings where the cost of false negatives or false positives can be significant.

TABLE 13. Detailed exploration of prior technologies with proposed biowordvec model.

Ref	Methodology	Applied Algorithm	Performance & Limitations
[12]	Transformer Models (BioBERT, R-BioBERT) for DDI Extraction.	Relation Extraction Models	High accuracy (~85%) but computationally expensive and requires extensive labeled datasets.
[13]	SMILES-Based Deep Learning Model for DDI.	Deep Neural Network (DNN)	Good accuracy (93.2%) but limited to molecular similarity, neglecting real-world pharmacovigilance data.
[17]	Graph Neural Networks (GNNs) for DDI Prediction.	Graph Attention Network (IPGAT)	High AUROC (~80%), but requires structured graphs, making it less adaptable to real-world datasets.
[21]	Deep Learning for Adverse Drug Interaction Prediction.	Deep Attributed Embedding Model (DAEM)	Captures high-dimensional relationships but needs extensive labeled data (AUC = 0.897).
[22]	Pharmacovigilance-Based DDI Prediction.	Siamese Neural Networks	Lacks domain-specific adaptation; general-purpose embedding struggle with rare drugs.
[23]	GNN with Hypergraph-Based DDI Learning.	CentSmoothie (Hypergraph Neural Network).	Captures complex interactions, but depends on structured datasets and is less effective for FAERS-type reports.
[24]	ChatGPT for DDI Prediction.	Large Language Model (LLM)	Limited accuracy and inconsistent reasoning for predictions.
[33]	AI-driven DDI prediction using multiple interaction models.	AI models integrating NLP and Knowledge Graphs	Provides comprehensive DDI prediction but requires high-quality data
[34]	Graph Neural Networks (GNNs) and Deep Learning for DDI prediction	GNNs and deep learning comparative study	Highly effective but computationally expensive
[35]	AI-driven DDI prediction using Knowledge Graphs and NLP	NLP, Deep Learning, and Knowledge Graphs	Scalable but requires large datasets
[36]	AI-based approaches for various DDI prediction types	Machine Learning, Deep Learning	Comprehensive but requires detailed model tuning
[37]	Comparative review of ML-based DDI prediction methods	Shallow Learning, Deep Learning, Knowledge Graph-based	Comprehensive review, highlights challenges and future directions
Proposed Model	BioWordVecIndicationSubstance Embedding + CatBoost Classifier	Fine-tuned word embedding with ML classification	Best F1 Score (73.32%), ROC AUC (84%).

- Comparative Analysis: The visualization clearly depicts the incremental improvements in model performance as more sophisticated embedding are employed. The transition from BioWordVec to BioWordVecIndication-Substance illustrates significant gains, demonstrating the value of enriching embedding with domain-specific knowledge.

The findings from this study contribute profoundly to the fields of pharmacovigilance and clinical informatics. By employing advanced ML techniques and sophisticated embedding methods, the prediction of the seriousness of drug-drug interactions is considerably enhanced.

These advancements deepen our understanding of DDIs and pave the way for integrating these analytical techniques into clinical practice, thus improving patient safety and therapeutic outcomes. Moreover, using enriched embedding and tailor-made ML models marks a pivotal advancement in improving drug safety through better predictive models. This approach allows for a nuanced analysis that can adapt to

the complexities of real-world clinical data, setting a new standard for pharmacovigilance analytics.

This study demonstrates a promising approach for predicting the seriousness of drug-drug interactions (DDIs) using ML and enriched semantic embedding. However, there are a few areas for potential improvement:

- Data Reporting Bias: The FAERS dataset is based on self-reported adverse events, which may result in some biases. Not all events are reported equally, and certain drug interactions may be more frequently documented than others.

Despite this, the dataset remains a valuable resource for pharmacovigilance, and our model's performance shows that useful patterns can still be identified.

- Binary Classification Approach: We utilized a binary classification to distinguish between serious and non-serious interactions. While this approach provides a straightforward prediction, future work could explore more detailed classifications to capture the varying levels of seriousness in drug interactions.

- Generalizability: The model was trained using FAERS data, which may have specific characteristics. While our results are promising, applying the model to other datasets or real-world clinical settings may require further validation to ensure its broader applicability.

These limitations provide directions for future research but do not detract from the demonstrated potential of our approach in improving drug safety analysis.

While this study primarily focuses on predicting the seriousness of drug-drug interactions using various embedding and ML techniques, we acknowledge the importance of understanding model performance across different contexts, such as drug classes, indications, and patient demographics. Most previous studies on DDI severity prediction rely on deep learning, graph-based models, or molecular representations, which are computationally inefficient, have limited real-world applicability, or are based on structured data. To address these challenges directly, we introduce domain-specific, fine-tuned embedding (BioWordVecIndicationSubstance), which captures richer contextual information from FAERS. As presented in Table 13, a detailed explanation of prior studies and technologies is given.

Due to the scope and structure of the current dataset, we did not explicitly stratify the training or testing processes based on drug classes, specific indications, or demographic groups. However, including demographic features (e.g., age and gender) during training allows the model to account for these variables implicitly in its predictions. In our experiments, adding demographic features had a limited impact on the overall model performance, suggesting that these factors might not strongly influence the prediction of DDI seriousness in this dataset.

For future work, we propose further experiments that involve separate modeling for different drug classes and explicit demographic stratification. This would help gain deeper insights into the nuances of drug interactions and how they may differ based on patient-specific factors or drug characteristics.

VIII. CONCLUSION

This study introduces a novel approach for predicting drug-drug interactions (DDIs) severity by integrating advanced semantic embedding and optimized machine-learning classification techniques.

The research addresses key limitations in traditional DDI prediction methods, including limited contextual feature representation, reliance on structured datasets, and computational inefficiencies. To overcome these challenges, we developed the BioWordVecIndicationSubstance embedding, which enhances semantic understanding by incorporating drug names, indications, and active substances, increasing embedding coverage from 40% to 80%. Our approach successfully captures deeper pharmacological relationships by leveraging transfer learning, improving model robustness in real-world applications. Our methodology evaluates multiple machine learning classifiers. CatBoost is the best performer

due to its ability to handle high-dimensional categorical data, optimize decision boundaries, and maintain strong classification performance. The experimental results demonstrate that our approach significantly improves prediction accuracy, with better F1 scores and ROC AUC than baseline models. The findings confirm that integrating semantic-rich embedding with advanced classification techniques enhances DDI severity prediction, addressing the initial research challenge of improving model reliability for pharmacovigilance.

In addition to methodological advances, this study emphasizes the practical implications for clinical decision-making and drug safety monitoring. The proposed model can be integrated into pharmacovigilance systems to help healthcare professionals detect potentially severe drug interactions and thus minimize adverse drug events. Nevertheless, our approach shows significant improvements, yet further work is required to validate its generalizability on multiple datasets other than FAERS and explore alternative embedding refinements for better interpretability. Future research can also incorporate biomedical knowledge graphs or deep learning architectures to refine DDI severity classification further.

Finally, this study demonstrates that a combination of enhanced semantic embedding and advanced machine learning can significantly improve the prediction of DDI severity. This work addresses the limitations of previous methods while maintaining computational efficiency and scalability. It also contributes to the ongoing development of data-driven pharmacovigilance strategies, leading to more accurate and actionable DDI risk assessments.

AUTHOR CONTRIBUTIONS

Conceptualization, Ayman Mohamed Mostafa, Alaa S. Alaerjan, Bader Aldughayfiq, Hisham Allahem, and Mohamed Ezz; methodology, Alshaimaa A. Tantawy, Mohamed Ezz, and Bader Aldughayfiq; data curation, Ayman Mohamed Mostafa, Bader Aldughayfiq, and Mohamed Ezz; formal analysis, Ayman Mohamed Mostafa and Mohamed Ezz; investigation, Alshaimaa A. Tantawy, Meshrif Alruily, and Bader Aldughayfiq; resources, Ayman Mohamed Mostafa and Mohamed Ezz; supervision, Bader Aldughayfiq, Alaa S. Alaerjan, Alshaimaa A. Tantawy, Hisham Allahem, and Mohamed Ezz; writing—original draft, A.A., Mohamed Ezz, and A.M.; writing—review and editing, Ayman Mohamed Mostafa, Mohamed Ezz, and Bader Aldughayfiq. All authors have read and agreed to the published version of the manuscript.

INFORMED CONSENT STATEMENT

Not applicable.

DATA AVAILABILITY STATEMENT

The experimental data and the simulation results that support the findings of this study are available at: FDA Adverse Event Reporting System, <https://open.fda.gov/data/faers/>.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] Y. Liang, “DDI-SSL: Drug–drug interaction prediction based on substructure signature learning,” *Appl. Sci.*, vol. 13, no. 19, p. 10750, Sep. 2023, doi: [10.3390/app131910750](https://doi.org/10.3390/app131910750).
- [2] F. A. M. Carizio, I. D. V. de Souza, T. Z. Oliveira, L. S. Silva, N. C. A. Rodrigues, M. O. B. Zanetti, F. R. Varallo, and L. R. Leira-Pereira, “Prediction of adverse drug reactions in geriatric patients admitted to intensive care units,” *Farmacia Hospitalaria*, vol. 48, no. 6, pp. 286–289, Nov. 2024, doi: [10.1016/j.farma.2024.03.004](https://doi.org/10.1016/j.farma.2024.03.004).
- [3] P. Lin, E. Berg, D. Wei, E. Liu, and A. Ko, “Improved prediction of adverse drug reactions for loperamide and pramipexole with in vitro secondary pharmacology profiling panels,” *J. Pharmacological Toxicological Methods*, vol. 111, Sep. 2021, Art. no. 106973, doi: [10.1016/j.vascn.2021.106973](https://doi.org/10.1016/j.vascn.2021.106973).
- [4] M. Alruily, A. Manaf Fazal, A. M. Mostafa, and M. Ezz, “Automated Arabic long-tweet classification using transfer learning with BERT,” *Appl. Sci.*, vol. 13, no. 6, p. 3482, Mar. 2023, doi: [10.3390/app13063482](https://doi.org/10.3390/app13063482).
- [5] B. Aldughayfiq, H. Allahem, A. M. Mostafa, M. Alnusayri, and M. Ezz, “Layer-weighted attention and ascending feature selection: An approach for seriousness level prediction using the FDA adverse event reporting system,” *Appl. Sci.*, vol. 14, no. 8, p. 3280, Apr. 2024, doi: [10.3390/app14083280](https://doi.org/10.3390/app14083280).
- [6] S. V. Iyer, R. Harpaz, P. LePendu, A. Bauer-Mehren, and N. H. Shah, “Mining clinical text for signals of adverse drug–drug interactions,” *J. Amer. Med. Inform. Assoc.*, vol. 21, no. 2, pp. 353–362, Mar. 2014, doi: [10.1136/amiajnl-2013-001612](https://doi.org/10.1136/amiajnl-2013-001612).
- [7] K. Pozsgai, G. Szics, A. Künig-Péter, O. Balázs, P. Vajda, L. Botz, and R. G. Vida, “Analysis of pharmacovigilance databases for spontaneous reports of adverse drug reactions related to substandard and falsified medical products: A descriptive study,” *Frontiers Pharmacol.*, vol. 13, Sep. 2022, Art. no. 964399, doi: [10.3389/fphar.2022.964399](https://doi.org/10.3389/fphar.2022.964399).
- [8] L. Wang, A. Shendre, C. Chiang, W. Cao, X. Ning, P. Zhang, P. Zhang, and L. Li, “A pharmacovigilance study of pharmacokinetic drug interactions using a translational informatics discovery approach,” *Brit. J. Clin. Pharmacol.*, vol. 88, no. 4, pp. 1471–1481, Apr. 2022, doi: [10.1111/bcpt.14762](https://doi.org/10.1111/bcpt.14762).
- [9] A. Ben Abacha, M. F. M. Chowdhury, A. Karanasiou, Y. Mrabet, A. Lavelli, and P. Zweigenbaum, “Text mining for pharmacovigilance: Using machine learning for drug name recognition and drug–drug interaction extraction and classification,” *J. Biomed. Informat.*, vol. 58, pp. 122–132, Dec. 2015, doi: [10.1016/j.jbi.2015.09.015](https://doi.org/10.1016/j.jbi.2015.09.015).
- [10] K. Han, P. Cao, Y. Wang, F. Xie, J. Ma, M. Yu, J. Wang, Y. Xu, Y. Zhang, and J. Wan, “A review of approaches for predicting drug–drug interactions based on machine learning,” *Frontiers Pharmacol.*, vol. 12, Jan. 2022, Art. no. 814858, doi: [10.3389/fphar.2021.814858](https://doi.org/10.3389/fphar.2021.814858).
- [11] C. Chukwuocha, T. Mathu, and K. Raimond, “Design of an interactive biomedical text mining framework to recognize real-time drug entities using machine learning algorithms,” *Proc. Comput. Sci.*, vol. 143, pp. 181–188, Jan. 2018, doi: [10.1016/j.procs.2018.10.374](https://doi.org/10.1016/j.procs.2018.10.374).
- [12] M. Kafikang and A. Hendawi, “Drug–drug interaction extraction from biomedical text using relation BioBERT with BLSTM,” *Mach. Learn. Knowl. Extraction*, vol. 5, no. 2, pp. 669–683, Jun. 2023, doi: [10.3390/make5020036](https://doi.org/10.3390/make5020036).
- [13] X. Hou, J. You, and P. Hu, *Predicting Drug–Drug Interactions Using Deep Neural Network*, 2019, pp. 168–172.
- [14] N. Rohani and C. Eslahchi, “Drug–drug interaction predicting by neural network using integrated similarity,” *Sci. Rep.*, vol. 9, no. 1, p. 13645, Sep. 2019, doi: [10.1038/s41598-019-50121-3](https://doi.org/10.1038/s41598-019-50121-3).
- [15] Y. Deng, X. Xu, Y. Qiu, J. Xia, W. Zhang, and S. Liu, “A multimodal deep learning framework for predicting drug–drug interaction events,” *Bioinformatics*, vol. 36, no. 15, pp. 4316–4322, Aug. 2020, doi: [10.1093/bioinformatics/btaa501](https://doi.org/10.1093/bioinformatics/btaa501).
- [16] S. Mei and K. Zhang, “A machine learning framework for predicting drug–drug interactions,” *Sci. Rep.*, vol. 11, no. 1, p. 17619, Sep. 2021, doi: [10.1038/s41598-021-97193-8](https://doi.org/10.1038/s41598-021-97193-8).
- [17] J. Wang, C. Guo, and X. Wu, “Predicting drug–drug interactions with graph attention network,” in *Proc. 26th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2022, pp. 4953–4959, doi: [10.1109/ICPR56361.2022.9956556](https://doi.org/10.1109/ICPR56361.2022.9956556).
- [18] Z. Liu, X.-N. Wang, H. Yu, J.-Y. Shi, and W.-M. Dong, “Predict multi-type drug–drug interactions in cold start scenario,” *BMC Bioinf.*, vol. 23, no. 1, p. 75, Dec. 2022, doi: [10.1186/s12859-022-04610-4](https://doi.org/10.1186/s12859-022-04610-4).
- [19] S. Yang and S. Kar, “Application of artificial intelligence and machine learning in early detection of adverse drug reactions (ADRs) and drug-induced toxicity,” *Artif. Intell. Chem.*, vol. 1, no. 2, Dec. 2023, Art. no. 100011, doi: [10.1016/j.aichem.2023.100011](https://doi.org/10.1016/j.aichem.2023.100011).
- [20] L. Zhuang, H. Wang, J. Zhao, and Y. Sun, “Adaptive dual graph contrastive learning based on heterogeneous signed network for predicting adverse drug reaction,” *Inf. Sci.*, vol. 642, Sep. 2023, Art. no. 119139, doi: [10.1016/j.ins.2023.119139](https://doi.org/10.1016/j.ins.2023.119139).
- [21] J. Zhu, Y. Liu, Y. Zhang, Z. Chen, K. She, and R. Tong, “DAEM: Deep attributed embedding based multi-task learning for predicting adverse drug–drug interaction,” *Exp. Syst. Appl.*, vol. 215, Apr. 2023, Art. no. 119312, doi: [10.1016/j.eswa.2022.119312](https://doi.org/10.1016/j.eswa.2022.119312).
- [22] P. Gupta, A. Majumdar, E. Chouzenoux, and G. Chierchia, “DeConDF-Fuse: Predicting drug–drug interaction using joint deep convolutional transform learning and decision forest fusion framework,” *Exp. Syst. Appl.*, vol. 227, Oct. 2023, Art. no. 120238, doi: [10.1016/j.eswa.2023.120238](https://doi.org/10.1016/j.eswa.2023.120238).
- [23] D. A. Nguyen, C. H. Nguyen, and H. Mamitsuka, “Central-smoothing hypergraph neural networks for predicting drug–drug interactions,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 8, pp. 11620–11625, Aug. 2024, doi: [10.1109/TNNLS.2023.3261860](https://doi.org/10.1109/TNNLS.2023.3261860).
- [24] A. Juhi, N. Pipil, S. Santra, S. Mondal, J. K. Behera, and H. Mondal, “The capability of ChatGPT in predicting and explaining common drug–drug interactions,” *Cureus*, vol. 15, no. 3, Mar. 2023, Art. no. e36272, doi: [10.7759/cureus.36272](https://doi.org/10.7759/cureus.36272).
- [25] J. Gao, Z. Wu, R. Al-Sabri, B. M. Oloulade, and J. Chen, “AutoDDI: Drug–drug interaction prediction with automated graph neural network,” *IEEE J. Biomed. Health Informat.*, vol. 28, no. 3, pp. 1773–1784, Mar. 2024, doi: [10.1109/JBHI.2024.3349570](https://doi.org/10.1109/JBHI.2024.3349570).
- [26] S. Chen, I. Semenov, F. Zhang, Y. Yang, J. Geng, X. Feng, Q. Meng, and K. Lei, “An effective framework for predicting drug–drug interactions based on molecular substructures and knowledge graph neural network,” *Comput. Biol. Med.*, vol. 169, Feb. 2024, Art. no. 107900, doi: [10.1016/j.combiomed.2023.107900](https://doi.org/10.1016/j.combiomed.2023.107900).
- [27] M. Dou, J. Tang, P. Tiwari, Y. Ding, and F. Guo, “Drug–drug interaction relation extraction based on deep learning: A review,” *ACM Comput. Surv.*, vol. 56, no. 6, pp. 1–33, Jun. 2024, doi: [10.1145/3645089](https://doi.org/10.1145/3645089).
- [28] D. Zaikis and I. Vlahavas, “Drug–drug interaction classification using attention based neural networks,” presented at the *Proc. 11th Hellenic Conf. Artif. Intell.*, Athens, Greece, Sep. 2020, doi: [10.1145/3411408.3411461](https://doi.org/10.1145/3411408.3411461).
- [29] F. Youbi and N. Settouti, “Analysis of machine learning and deep learning frameworks for opinion mining on drug reviews,” *Comput. J.*, vol. 65, no. 9, pp. 2470–2483, Sep. 2022, doi: [10.1093/comjnl/bxb084](https://doi.org/10.1093/comjnl/bxb084).
- [30] Y. Zhang, Q. Chen, Z. Yang, H. Lin, and Z. Lu, “BioWordVec, improving biomedical word embeddings with subword information and MeSH,” *Sci. Data*, vol. 6, no. 1, p. 52, May 2019, doi: [10.1038/s41597-019-0055-0](https://doi.org/10.1038/s41597-019-0055-0).
- [31] B. Kompa, J. B. Hakim, A. Palepu, K. G. Kompa, M. Smith, P. A. Bain, S. Woloszynek, J. L. Painter, A. Bate, and A. L. Beam, “Artificial intelligence based on machine learning in pharmacovigilance: A scoping review,” *Drug Saf.*, vol. 45, no. 5, pp. 477–491, May 2022, doi: [10.1007/s40264-022-01176-1](https://doi.org/10.1007/s40264-022-01176-1).
- [32] *FDA Adverse Event Reporting System*. [Online]. Available: <https://open.fda.gov/data/faers/>
- [33] S. Chen, T. Li, L. Yang, F. Zhai, X. Jiang, R. Xiang, and G. Ling, “Artificial intelligence-driven prediction of multiple drug interactions,” *Briefings Bioinf.*, vol. 23, no. 6, Nov. 2022, Art. no. bbac427, doi: [10.1093/bib/bbac427](https://doi.org/10.1093/bib/bbac427).
- [34] X. Lin, L. Dai, Y. Zhou, Z.-G. Yu, W. Zhang, J.-Y. Shi, D.-S. Cao, L. Zeng, H. Chen, B. Song, P. S. Yu, and X. Zeng, “Comprehensive evaluation of deep and graph learning on drug–drug interactions prediction,” *Briefings Bioinf.*, vol. 24, no. 4, Jul. 2023, Art. no. bbad235, doi: [10.1093/bib/bbad235](https://doi.org/10.1093/bib/bbad235).
- [35] P. Sonaji, L. Subramanian, and M. Rajesh, “Artificial intelligence-driven drug interaction prediction,” *World J. Biol. Pharmacy Health Sci.*, vol. 17, no. 2, pp. 297–305, Feb. 2024, doi: [10.30574/wjbphs.2024.17.2.0070](https://doi.org/10.30574/wjbphs.2024.17.2.0070).
- [36] Y. Zhang, Z. Deng, X. Xu, Y. Feng, and S. Junliang, “Application of artificial intelligence in drug–drug interactions prediction: A review,” *J. Chem. Inf. Model.*, vol. 64, no. 7, pp. 2158–2173, Apr. 2024, doi: [10.1021/acs.jcim.3c00582](https://doi.org/10.1021/acs.jcim.3c00582).

- [37] N.-N. Wang, B. Zhu, X.-L. Li, S. Liu, J.-Y. Shi, and D.-S. Cao, "Comprehensive review of drug–drug interaction prediction based on machine learning: Current status, challenges, and opportunities," *J. Chem. Inf. Model.*, vol. 64, no. 1, pp. 96–109, Jan. 2024, doi: 10.1021/acs.jcim.3c01304.



AYMAN MOHAMED MOSTAFA (Member, IEEE) received the M.Sc. and Ph.D. degrees in information systems from the Faculty of Computers and Informatics, Zagazig University, Egypt. He is currently an Associate Professor with the Faculty of Computers and Informatics, Zagazig University, and an Assistant Professor with the College of Computer and Information Sciences, Jouf University, Saudi Arabia. He has published more than 60 scientific papers in various national and international journals and conferences. His area of research interests include information security, cloud computing, e-business, e-commerce, big data, and data science.



ALAA S. ALAERJAN is currently the Vice-Rector for Academic Affairs of Jouf University, Saudi Arabia. He is an Associate Professor with the College of Computer and Information Sciences, Jouf University. His research interests include machine learning and artificial intelligence in biomedical applications.



HISHAM ALLAHEM received the Ph.D. degree in computer science from Dalhousie University, in 2022. He has been with Jouf University as an Assistant Professor, since 2022. He has a total of 16 years of experience in academic work. He has published papers in conferences and journals in areas of healthcare applications and technologies, wireless network security, machine learning, deep learning, and body sensor networks.



BADER ALDUGHAYFIQ received the master's and Ph.D. degrees in computer science from Dalhousie University, Halifax, Canada. He is currently an Assistant Professor with the Information Systems Department, Jouf University. His primary research interests include artificial intelligence, machine learning, and deep learning applications.



MESHRIIF ALRUILY received the Ph.D. degree in computer science from the University of De Montfort, U.K., in 2012. He is currently a Professor with the Department of Computer and Information Sciences, Jouf University, Saudi Arabia. He published many conference papers and journal articles. His research interests include natural language processing, social media analysis, the IoT, machine learning, and deep learning.

ALSHAIMAA A. TANTAWY, photograph and biography not available at the time of publication.



MOHAMED EZZ (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in systems and computer engineering from the Faculty of Engineering, Al Azhar University. He is currently an Associate Professor with the Faculty of Engineering, Al Azhar University, and a Visiting Professor with the College of Computer and Information Sciences, Jouf University. He has published 40 scientific papers in various national and international journals and conferences. He has contributed to more than 16 mega software projects in electronic banking EBPP, EMV, mobile banking, and e-commerce, also CBAP Certified. His area of research interests include pattern recognition, applied machine learning, application security, intrusion detection, and semantic web.