Journal of Posthumanism
2025
Volume: 5, No: 6, pp. 4414-4438
ISSN: 2634-3576 (Print) | ISSN 2634-3584 (Online)
posthumanism.co.uk

DOI: https://doi.org/10.63332/joph.v5i6.2635

A Proposed Model for Detecting Drug-Drug Interactions in Hypertension and Diabetes Treatments

Fatma M. Seoudi[1], Ahmed S. Ismail[2], Muhammed H. Ebrahim[3]

Abstract

Hypertension and diabetes are two of the most prevalent chronic diseases in the world, and in their management, a drug combination is usually required. However, drug-drug interactions (DDIs) cause harmful side effects, endangering the safety of the patient and the efficacy of treatment. The proposed study examines the issues of co-management of medications for hypertension and diabetes. The mechanism of action and metabolic pathways common to these drugs create a complex environment, increasing the risk of side effects and decreasing the efficacy of treatment. Artificial intelligence (AI) can be employed to predict drug interactions, pharmacodynamics, and side effects. In this research, an intelligent model is implemented to forecast the outcome of drug interactions in the treatment of diabetes and hypertension. The proposed method can assist in reducing the time and expense required to identify the optimal drug combinations for clinical applications. The DDI dataset was obtained from the Drug Bank database, taking into account drug pairs for hypertension and diabetes. Additionally, some chemical features are extracted from the Simplified Molecular Input Line Entry System (SMILES) strings of the drug pairs. Different machine learning algorithms are applied to these features, such as Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), and eXtreme Gradient Boosting (XGBoost). the performance of XGBoost was better than the others and achieved outstanding accuracy. The proposed model predicted a mean of 97% accuracy for DDIs. The study again highlights the importance of computationally founded predictive modeling in promoting the efficiency of drug administration and lowering the incidence of adverse reactions.

Keywords: *Hypertension, Diabetes, Drug-Drug Interactions, Artificial Intelligence, Simplified Molecular Input Line Entry System, PyBioMed, Drug Bank, Machine Learning.*

Introduction

The cardiovascular and endocrine systems play a crucial role in maintaining homeostasis in the human body. Two of the most chronic conditions globally are diabetes and hypertension, which both significantly increase the risk of developing devastating complications such as cardiovascular disease, renal failure, and neuropathy. Treatment of these conditions is commonly accomplished by polypharmacy, the use of more than one medication simultaneously, which raises concern about drug-drug interactions (DDIs). DDIs occur when two or more drugs interact with each other's pharmacological effect, either enhancing, diminishing, or altering their therapeutic and safety profiles (Kuhn et al., 2010; Lounkine et al., 2012; Saverno et al., 2011). In treating comorbid diabetes and hypertension, combination treatment using antidiabetic and antihypertensive drugs is usually required. Although such drugs may differ in their mechanisms of action and metabolism pathways, their concurrent interactions may lead to significant DDIs. These can interfere with blood pressure regulation, exacerbate glycemic control, or cause adverse effects such as hypotension, hypoglycemia, electrolyte imbalances, and other adverse complications (Kuhn et al., 2010;

[1] Faculty of Computers and Artificial Intelligence, Fayoum University, Fayoum 63514, Egypt, Faculty of Computers and Information Technology, National Egyptian E-learning University, Giza 12611, Egypt, Email: fm1489@fayoum.edu.eg, (Corresponding Author)
[2] Faculty of Computers and Artificial Intelligence, Fayoum University, Fayoum 63514, Egypt
[3] Faculty of Computers and Artificial Intelligence, Fayoum University, Fayoum 63514, Egypt

Lounkine et al., 2012; Saverno et al., 2011). A complete understanding of potential DDIs is therefore imperative in ensuring safe and effective treatment and necessitates careful screening of a patient's drug profile and any past adverse drug reactions (Lounkine et al., 2012). Despite the availability of clinical guidelines, DDIs prediction and prevention remains an arduous task due to the complexity of drug mechanisms and the tremendous volume of pharmacological data. Traditional approaches to DDI detection such as manual screening, are not just labor-intensive and time-consuming; they are also prone to human error. Moreover, most drug interactions are not realized during preclinical stages and appear only during clinical trials or post-marketing surveillance, resulting in serious health and economic costs (Abacha et al., 2015; Cheng & Zhao, 2014; Ferdousi et al., 2017; Percha & Altman, 2013; Saverno et al., 2011).

Over the past decade, computational methods, particularly AI- and ML-based methods, have shown promise in DDI prediction more efficiently and accurately (Cheng & Zhao, 2014; Percha & Altman, 2013). AI algorithms can automatically process large pharmacological databases and extract potential interactions from chemical structure, pharmacokinetics, and clinical data. Databases such as DrugBank (Wishart et al., 2018), DailyMed (Breslow et al., 2015), Human Metabolome Database) Wishart et al., 2018(, and Drug Interaction Facts (Tatro, 2011)provide a good foundation for training predictive machine learning models. Several works have applied ML and deep learning to predict DDIs. For example, Gottlieb et al. (Gottlieb et al., 2012) proposed the "Inferring Drug Interactions" method derived from pharmacodynamic and metabolic data. You et al. (et al. 2019) integrated drug-target interaction datasets and applied LASSO regression models with high prediction accuracy. Lee et al.) Lee et al., 2019(developed a deep feed-forward neural network that integrated structure similarity, Gene Ontology (GO), and target gene similarity (TSP) in 2019 and attained 96% prediction accuracy. More recently, Deng et al. (Deng et al., 2020) introduced DDIMDL, a multimodal deep learning model that integrates chemical structures, enzymes, and biological pathways, which also exhibited excellent performance. But the majority of these models are plagued by imbalanced datasets and sparse interaction data, leading to underfitting and lower generalizability. Predictive performance improvement through data augmentation and more sophisticated feature engineering remains an important area of research.

In this study, we propose a machine learning-based model for predicting potential DDIs between antidiabetic and antihypertensive drugs. We hypothesize that many DDIs result from interactions at the chemical level between functional groups of drug molecules with major metabolic enzymes, i.e., cytochrome P450 isoenzymes (Lynch & Price, 2007). By extracting features from drug pairs using their SMILES (Simplified Molecular Input Line Entry System) representations and applying state-of-the-art machine learning algorithms, we aim to build a predictive model with high capacity to distinguish between DDI and non-DDI drug pairs. With the increasing reliance on combination therapy in hypertension and diabetes treatment, especially in the elderly, who are most susceptible to ADEs, our approach can make a significant difference in patient safety, therapeutic efficacy, and cost and time savings in DDI discovery in drug development and clinical practice. The remaining sections of this paper are structured as follows: Section 2 covers related work on computational approaches in predicting drug-drug interaction (DDI) with special interest in antidiabetic and antihypertensive drugs. Section 3 describes the dataset, including explanations of drug pair selection, SMILES representation, and feature extraction processes. Section 4 gives the machine learning models and measures. Section 5 presents and discusses the experimental results and their significance. Section 6 concludes the paper with a summary of findings and recommendations for further research.

This study aims to evaluate the effects of foreign direct investment, trade openness, technological innovation, and industrial sector performance on economic growth in Vietnam. It seeks to provide empirical evidence that supports modern economic growth theory. The research methodology involves using time series data and Autoregressive Distributed Lag (ARDL) estimation techniques to analyze both the short-run and long-run impacts of these factors on economic growth. Additionally, the findings offer significant insights for shaping development policies that align with Vietnam's conditions and strategic goals in this new phase of development.

The structure of the study is organized as follows: Section 1: Introduction of the research problem; Section 2: Overview of related studies; Section 3: Description of the data sources and research methods; Section 4: Presentation of the empirical research results; Section 5: Conclusions and policy implications.

Literature Review

Through the research (Hung et al., 2022), an AI-based predictive model of drug-drug interactions for osteoporosis and Paget's disease from SMILES representations is proposed. The primary goal is to rationalize drug combinations to achieve better therapeutic effects and prevent unwanted drug interactions. The data set, which is collected from the Drug Bank database, is for osteoporosis and Paget's disease (OPD) drugs, and SMOTE is applied to deal with data imbalance.

Some machine learning algorithms were employed, including SVM, k-NN, Decision Trees (DT), Random Forest (RF), XGBoost, and stacked models to enhance predictive performance. Feature selection and model interpretability were achieved using tools like Weka and SHAP, enabling improved insight into the decision-making processes of the model. Experimental results demonstrated that the combined RF-XGBoost model had an accuracy level of 74%, which shows its effectiveness at predicting drug interactions. The findings demonstrate that the proposed model can potentially enhance clinical decision-making and reduce time and costs in healthcare management.

The authors in (Wu et al., 2020) were concerned with the extraction of drug-drug interaction (DDI) from biomedical literature with a hybrid neural network model. The proposed model integrates a couple of deep learning components, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs or BiLSTM), to effectively extract the linguistic and semantic context of biomedical text and accurately identify drug interaction relations.

The model scans sentences in scientific journals to determine if they describe a drug interaction, employing advanced natural language processing (NLP) techniques. Experimental results showed that the hybrid model achieved an F1 score of 0.75, which outperformed traditional and single-method approaches by exhibiting a high balance between recall and precision. The findings demonstrate the model's ability to inform intelligent DDI extraction systems to achieve improved drug safety and reduce clinical risk.

The work in (Zhu et al., 2020) introduces extracting drug-drug interactions (DDIs) from texts using BioBERT with various entity-aware attention mechanisms. The model uses the pre-trained BioBERT language model, specifically tailored for biomedical text, and layers of attention that are drug entity-aware so that it can attend more to drug contextual relationships within sentences. Through the processing of biomedical sentences, the model identifies the occurrence of a DDI,

with the benefit of deep contextualized language representations and entity-specific attention. The experiment results indicate that the model achieved an F1-score of 0.809, which outperformed existing state-of-the-art techniques on precision-recall balance. These results highlight the effectiveness of the model in optimizing DDI extraction techniques and the potential for applying it to enhance clinical decision support systems and drug safety.

On the other hand, the AGCN is presented by the authors in (Park et al., 2020) as a new drug-drug interaction (DDI) extraction model. The model exploits the representational capability of graph-structured and attention-based models in capturing the subtle relationships among drug entities in biomedical text accurately.

AGCN constructs graph representations of sentences, depicting entities and their relationships as nodes and edges. The graphs are subsequently subjected to graph convolutional networks (GCNs) and complemented with attention layers for improving focus on critical relational information. The model was capable of achieving an F1-score of 0.7686, which was robust relative to baseline approaches. The findings demonstrate the potential for graph-based attention frameworks to improve DDI extraction and biomedical text analysis.

Also, the study in (Allahgholi et al., 2020) suggests ADDI, an innovative system that is not just focused on recognizing harmful drug-drug interactions (DDIs), but also on proposing alternative medications that do not have negative health effects. Differing from current DDI extraction techniques, ADDI aims at therapeutic decision-making by offering potential drug replacements.

The model leverages existing DDI knowledge with machine learning to identify potentially harmful drug pairs and propose alternative drugs that do not have similar harmful interactions. ADDI achieves high accuracy in detection and recommendation at 0.954, and the model's performance on the two tasks is highlighted as being good. The results above confirm that the model can facilitate intelligent clinical decision-making and prevent toxic DDIs for improved patient safety.

In addition, this article (Fatehifar & Karshenas, 2021) presents a robust approach for drug-drug interaction (DDI) extraction from a position and similarity fusion-based attention mechanism. The novel model enhances drug relation prediction using positional features (sentence entity locations) and contextual similarity, and allows the attention mechanism to focus on more appropriate sections of the text.

Through the use of combined representations, the model channels the attention layers toward more substantial relational indications in biomedical sentences. The method achieved an F1-score of 0.783, which is a good balance of recall and precision, and also its superior performance compared to several traditional DDI extraction approaches. The study denotes the importance of employing positional and linguistic cues in improving the accuracy of extracting DDIs from biomedical literature.

Proposed Methodology

Our methodology includes different sub-processes as shown in FIGURE 1 and is described in the following sections.

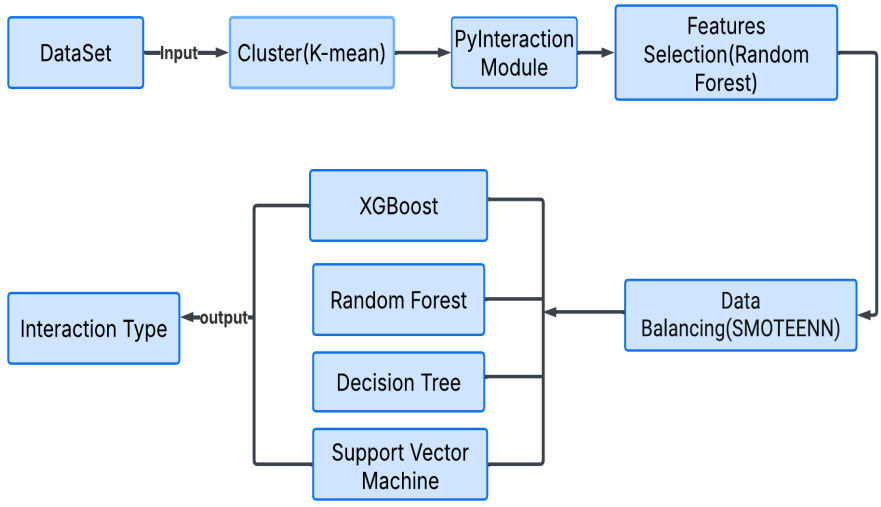

Figure 1: The Illustration of the Ddi Detection Framework Within The Hypertension and Diabetes Diseases

Data Preparation

Through this study, the DDI dataset was collected from the Drug Bank database (version 5.1.13, released on January 2, 2025) (*DrugBank*, 2025)within the hypertension and diabetes category, including only the drug pairs with the "approved" status (approved by the FDA) to ensure accuracy, while drug pairs labeled as "withdrawn," "experimental," or "investigated" were excluded from the dataset. Also, the DDI types are classified based on the specific description of the interaction (also collected from the Drug Bank database) between two drugs. For example, if an interaction is described as: "DRUG_A can cause a decrease in the absorption of DRUG_B, resulting in a reduced serum concentration and potentially a decrease in efficacy," where DRUG_A belongs to the hypertension or diabetes drug category, and DRUG_B is any drug that, when combined with DRUG_A, can lead to reduced serum concentration or decreased efficacy. The interactions were categorized into 13 types, ranging from class 0 to class 12, respectively. Hence, around 24,740 drug pairs are collected along with their detailed interaction descriptions and Simplified Molecular Input Line Entry System (SMILES) representations.

Feature Extraction

In this research, we hypothesize that drugs with chemical similarities are likely to exhibit similar biological activities, allowing us to identify interactions between drug pairs using a supervised dataset. The chemical similarities (features) were extracted from the SMILES representations of the defined hypertension and diabetes drug pairs using the PyInteraction module from the PyBioMed package. This package, written in Python, is designed to generate feature vectors from molecular structures, protein sequences, and DNA sequences. It is a versatile tool applicable to a wide range of tasks in fields related to cheminformatics, bioinformatics, and systems biology. The PyBioMed package consists of six main modules: PyInteraction, PyDNA, PyMolecule, PyProtein, PyGetMol, and PyPretreat. These modules compute various molecular descriptors and assist in processing input data. Within the PyInteraction module, there are three primary methods for calculating interaction descriptors, as implemented by the following functions.

$$F = \{(k) = F_{a(i)} * F_{\beta(j)} \mid i = 1, 2, \ldots, p\ ;\ j = 1, 2, \ldots, p\ ;$$

$$k = (i - 1) * p + j\} \qquad (1)$$

$$F_{a\beta} = (F_a, F_\beta) \qquad (2)$$

$$F = [F_{a(i)} + F_{\beta(i)}, F_{a(i)} * F_{\beta(i)}] \qquad (3)$$

The interaction features were created by multiplying the features of chemical *a* and chemical *b* in equation (1). In equation (2), the interaction descriptor was calculated by combining features of the two chemicals. Meanwhile, equation (3) can only be used for the same type of descriptor, such as chemical-chemical, protein-protein, and DNA-DNA interaction. In this study, equation (1) was chosen to generate the interaction descriptor of a pair of drugs because it produced the most features (3600) while the others only generated 120 features per pair of interactions.

The Interaction descriptors in the PyInteraction module are a combination of molecular properties in two-dimensional (2D) as shown in TABLE I such as Electro topological State Indices (Estate), Molecule topological polar surface area (MTPSA), Partition coefficients (log P), Molar refractivity (MR), and Van der Waals surface area (VSA), etc. As reported by Lo et al., 2D chemical descriptors are most frequently used in studies due to the invariant graph where descriptor values are not affected by the renumbering of graph nodes (vertices), making them useful for differentiation of special chemical structures.

Dimension	Interaction descriptor	Annotation
2D	MTPSA	Topological polar surface area based on molecules
2D	LabuteASA	La Bute's Approximate Surface Area
2D	LogPVSA	MOE-type descriptors using partition coefficients (log P) contributions and surface area contributions
2D	MRVSA	MOE-type descriptors using molar refractivity (MR) contributions and surface area contributions
2D	PEOEVSA	MOE-type descriptors using partial charges and surface area contributions
2D	EstateVSA	MOE-type descriptors using Estate indices and surface area contributions
2D	VSAEstate	MOE-type descriptors using surface area contributions and Estate indices

Table 1. Summary of Interaction Descriptors Using the Interaction Module for Feature Extraction

2D Molecular Interaction Descriptors:

For constructing robust predictive models of drug-drug interactions (DDIs), useful chemical features that describe the structural and electronic properties of molecules need to be extracted and represent the characteristics of molecules. In this work, a number of two-dimensional (2D) molecular interaction descriptors were utilized. These descriptors characterize a number of physicochemical and topological properties influencing drug behavior and interactions. Short descriptions of each descriptor utilized follow:

MTPSA (Topological Polar Surface Area):

This descriptor relies solely on the topological structure of the molecule to estimate its polar surface area, as opposed to 3D geometry. It provides an approximation of a molecule's hydrogen-bonding ability and is often employed to predict absorption, permeability, and penetration across the blood-brain barrier.

LabuteASA (Labute's Approximate Surface Area):

An atomic contribution-based estimation of molecular surface area, a process which was first devised by Labute. It is a quick estimation of the accessible surface area of the molecule, which finds utility in estimating molecular interactions and bioavailability.

LogPVSA (LogP-based Van der Waals Surface Area):

This descriptor combines Van der Waals surface area contributions and partition coefficient values (logP). It incorporates both hydrophobicity and surface exposure of the molecule and gives information regarding solubility and membrane permeability.

MRVSA (Molar Refractivity-based Van der Waals Surface Area):

This descriptor is a mix of molar refractivity (a measure of polarizability and molecular volume) and Van der Waals surface area. It can be used for explaining how electronic effects and steric interactions control molecular binding and recognition.

PEOEVSA (Partial Equalization of Orbital Electronegativities VSA):

Based on partial charges computed from orbital electronegativities, this descriptor adds surface area information to describe the electronic distribution in a molecule. It is especially effective for electrostatic interaction modeling with targets such as enzymes and receptors.

EstateVSA (Electro topological State Index-based Van der Waals Surface Area):

This two-descriptor uses electro topological state (E-state) indices and Van der Waals surface area. It considers the influence of the electronic environment as well as topology on molecular activity, making it suitable for quantitative structure-activity relationship (QSAR) modeling.

VSAEstate (Van der Waals Surface Area + E-State Indices):

Similar to EstateVSA, this descriptor is based on Van der Waals surface area and incorporates E-state indices. It is used to quantify the steric/electronic balance of molecular interactions.

Feature Selection

Feature importance refers to the process of assigning weights to the input features of the predictive model to compute the relative feature importance in terms of prediction. In the present study, we employed the Random Forest approach and Recursive Feature Elimination with Cross-Validation (RFECV) to identify the most important features from the provided dataset concerning drug interactions in hypertensive and diabetic therapy. The main goal is to select the most informative features to enhance model accuracy and efficiency with reduced computational complexity.

Firstly, the dataset is loaded and includes various drug-pair data and their corresponding interaction classes. For effective feature selection, non-numeric columns such as ***'smiles_a'***, ***'smiles_b', 'drug_a'***, and ***'drug_b***' were excluded from the feature set (X), while the 'Cluster'

column was designated as the target variable (y).

Feature Selection through Random Forest and RFECV:

We applied the Random Forest algorithm, which is run using the Random Forest Classifier from sklearn. Ensemble. We chose this model because of its ability to handle high-dimensional data and provide feature importance scores. To perform feature selection automatically, we utilized RFECV with the step parameter at 0.1 and 2-fold cross-validation to progressively drop the least significant features and to optimize model performance based on accuracy.

Reducing Feature Subset to 500 Features:

After the initial step of RFECV, whenever the number of features that had been identified exceeded 500, we again decreased the feature subset by re-training the Random Forest classifier with the picked features. We sorted these features according to their importance and considered only the most important 500 features having larger relevance scores. This operation provides a balance between model performance and computational burden.

Final Dataset Construction:

The final dataset was reconstructed in a way that incorporates the required identifiers ('smiles_a', 'smiles_b', 'drug_a', 'drug_b', and 'Cluster') as well as the selected features. This enriched dataset was saved as a CSV file for additional scrutiny and model construction. By this organized feature selection, we reduced the dimension of the data set with maximum information preserved to make sure that our model improved with a better ability to predict drug-drug interactions accurately, as far as hypertensive as well as diabetic conditions were concerned.

Machine Learning Model Development

The present research has entailed the use of several algorithms that include Support Vector Machine (SVM Decision Tree (DT), Random Forest (RF), and eXtreme Gradient Boosting (XGBoost).

1) Hyperparameter Optimization Using Randomized Search:

First, when applying the XGBoost model with its default hyperparameters, the accuracy was very poor at just 80%. This poor performance indicated that the initial configuration of the model was not very effectively picking up on the complex patterns of the dataset and thus was making very bad predictions. To address this problem, we used Randomized Search Cross-Validation (Randomized Search CV) to obtain the optimal hyperparameters for the XGBoost model. It allows for efficient exploration of a large search space using random selection from specified hyperparameter ranges. Upon performing hyperparameter tuning, test accuracy for the XGBoost model was greatly improved from 63% to 97.86%, thus showing the inherent role played by hyperparameter tuning in optimizing the performance of a model.

Similarly, Randomized Search CV was also employed in the other models (SVM, Random Forest, and Decision Tree), and their predictive accuracy increased greatly with significant improvements. This emphasizes how hyperparameter tuning is important in improving machine learning model performance, especially for complex tasks such as drug-drug interaction (DDI) prediction.

The process of hyperparameter optimization with the Randomized Search approach (Randomized Search CV) was performed on the entire training dataset. It gave back a specified

subset of parameters that were suitable for each machine-learning model, as shown in TABLE 2.

Algorithms	Hyperparameter	Randomized Search Range	Optimal value
Support Vector Machine	C	0.001,0.01,0.1,1,10	1
	Gamma	0.01, 0.1, 0.2, 0.3	0.01
	Kernel	RBF, linear	Rbf
Decision Tree	Criterion	gini, entropy	Entropy
	max_depth	10, 15, 20,25	15
	min_samples_leaf	2, 11, 2	4
	min_samples_split	5, 21 5	5
Random Forest	max_depth	10, 15, 20	15
	min_samples_leaf	2, 5	2
	min_samples_split	5,10	5
	n_estimators	50, 100, 150	100
XGBoost	max_depth	3, 6, 10	6
	n-estimators	50, 100, 200	200
	Learning_rate	0.01, 0.1, 0.2	0.2
	Colsample_bytree Subsample	0.7, 0.8, 0.9 0.7, 0.8, 0.9	0.8 0.9

Table 2. Randomized Search for Hyper Parameters and Optimal Value for Each Machine Learning Model

1. Support Vector Machine:

The main principle of SVM *(Vapnik & Izmailov, 2021)*is to divide the different classes from each other by constructing an optimal hyperplane or set of hyperplanes in high-dimensional space. The training points closest to this margin are the named support vector, thus giving the algorithm its name.

Given a labeled dataset:

$(x_1, y_1), \ldots, (x_n, y_n),\ x_i \in \mathbb{R}^d$ *and* $y_i \in$ *(Abacha et al., 2015)* (4)

In which x_i is the features used as the inputs for the model, and y_i is the target. The optimal hyperplane in this case will be:

$$w \cdot x + b = 0 \quad (5)$$

2. Decision Tree

DT *(Farid et al., 2014)* is a predictive model, constructed by analyzing a set of training examples for which the class labels are known. It classifies the data by generating a series of questions about the attributes associated with the target. In this study, the interaction types or Y are denoted

(13 classes - dependent variables) as the target variables. X1, X2, X3, and so on are variables equivalent to descriptors that act as input data to contribute to the decision of the type of interaction between the given drug pair.

3. Random Forest:

RF or random decision forest was first introduced by Tin Kam Ho in 1998.(Ho & intelligence, 1998) Ho(Ho & intelligence, 1998)invented the random subspace method for constructing ensemble models. The term Random Forest was coined by Leo Breiman (Breiman, 2001) in 2001 and it includes combining the random subspace method by Ho and bagging.(Breiman, 1996). It is an ensemble learning algorithm, that is, it amalgamates the results of several base models, in this case, decision trees, to produce an optimum model.

4. XGBoost:

XGBoost is an optimized distributed gradient boosting classifier designed to be highly efficient, flexible, and portable.(Chen & Guestrin, 2016) XGB is normally used for supervised learning problems, in this study, the XGB used the training data-x_i (with multiple features extracted from the PyInteraction module) to predict various targets y_i (multiple DDI types).

5. Combating Imbalance Classes:

Class imbalance is a common problem in machine learning classification problems, where the sample count is extremely dissimilar across classes. Class imbalance may result in predictive models favoring the majority classes, reducing accuracy when predicting minority classes.

Hence, the SMOTEENN technique was employed, using two methods as the following:

SMOTE (Synthetic Minority Oversampling Technique): This method generates synthetic samples for the minority classes, thereby increasing their count.

ENN (Edited Nearest Neighbors): ENN removes noisy and uncertain samples of the majority classes and hence improves the dataset quality.

As shown in Figure 2, before applying SMOTEENN, the distribution of the number of samples between classes was very uneven with some classes overrepresented and others underrepresented. After applying the technique, a more balanced distribution was achieved where both minority classes were oversampled and noisy data from the majority classes were eliminated as shown in FIGURE 3. The default parameters of SMOTEENN are used as they were observed to work well in balancing the dataset, improving the model's performance, and its ability to predict correctly underrepresented classes.

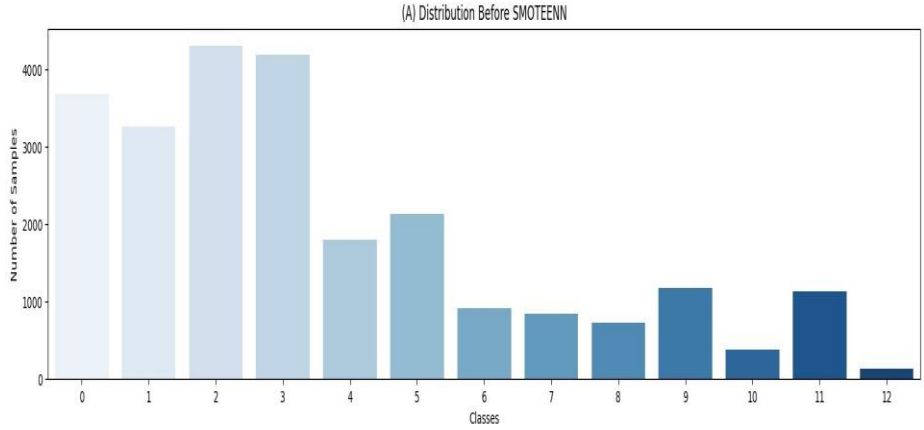

Figure 2. The sample distribution before applying the SMOTEENN method indicated a large class imbalance.

The sample distribution before applying the SMOTEENN method indicated a large class imbalance, with some classes being grossly overrepresented while others were significantly underrepresented. This imbalance presented a challenge in model training, as it had the potential to result in biased predictions biased towards the majority classes and decreased sensitivity towards minority class interactions.

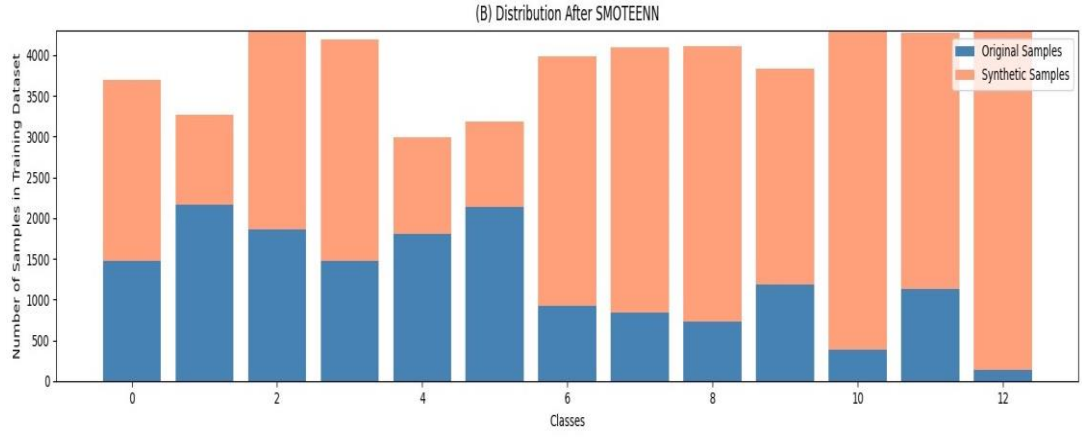

Figure 3.

After applying the SMOTEENN algorithm, the class distribution in the dataset became more balanced

After applying the SMOTEENN algorithm, the class distribution in the dataset became more balanced. The minority classes were effectively oversampled for improved representation, while noisy or misclassified instances in the majority classes were removed. Balancing made the dataset of improved quality and enabled the model's ability to generalize across all the classes.

Implementation and Discussion:

To calculate feature importance and observe how every variable is helping in the model's

predictions, we employed the XGBoost algorithm—a powerful tree-based machine learning algorithm with high performance and handling imbalanced data. The model outputs were interpreted with Shapley Additive explanations (SHAP), a complex technique that provides accurate explanations of how every independent variable influences the model's predictions.

After cleaning and balancing the dataset using the SMOTEENN technique, the data was split into a training and test set. The XGBoost model was trained using the optimal hyperparameters identified using Randomized Search CV.

For feature contribution estimation, a SHAP Explainer was initialized to estimate the contribution of each feature to the model's output. Results were plotted as follows:

- Bar Plot: Orders the features by importance, with the most influential features at the top, as shown in FIGURE 4. This analysis provides a direct idea about the most important features that are affecting the model's decisions, which makes the results more interpretable and credible.

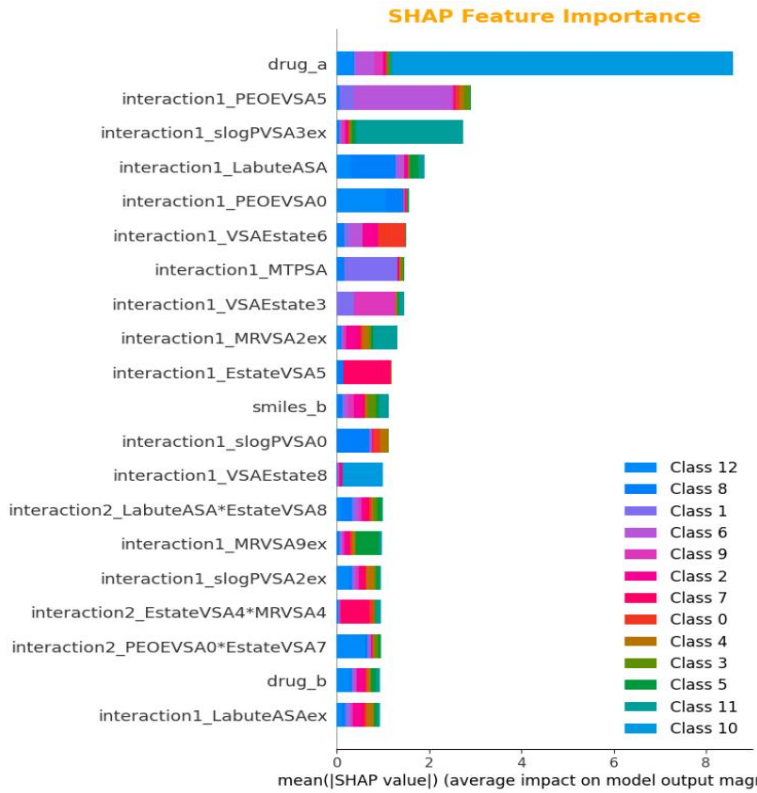

Figure 4. SHAP Features Plot.

SHAP Features Plot. The SHAP Features Plot shows the SHAP values for a specific feature across all instances, allowing us to understand how the feature is influencing the model's prediction for each sample. The plot reveals whether higher or lower values of the feature correspond to higher or lower predictions. It also provides insights into interactions with other features and potential patterns in the model's behavior.

Results

A proposed model is developed and experimented with machine learning methods using a Drug-Drug Interaction (DDI) dataset we downloaded from the Drug Bank database in this study. The dataset is aimed at adverse drug interactions only. The experimental setup is structured into three primary steps:

Data Collection and Filtering:

Pertinent drug data was collected, such as drug IDs, SMILES representations, and elaborate descriptions of drug interactions. The analysis focused on drugs in a given group and those that interact with them.

Feature Extraction and Labeling:

The interaction types were labeled, and important features extracted from pairs of drugs were obtained through SMILES descriptions and the PyInteraction module. The process produced a large set of molecular descriptors that reflected the chemical characteristics of each pair of drugs for subsequent use.

Model Training and Evaluation:

To solve the class imbalance issue, the SMOTEENN algorithm was utilized to balance the data. Testing and training of the machine learning model were done according to different classification algorithms. 5-fold cross-validation (5-fold CV) was utilized for performance evaluation, in which the dataset was randomly divided into five equal groups. Four subsets were used in every iteration to train, and one subset to validate.

Hence, the model performance was calculated via the following performance metrics:

Precision: The fraction of correctly predicted DDIs over all positive predictions.

Formula: $\mathbf{Precision} = \frac{\mathbf{TruePositives}}{\mathbf{TruePositives+FalsePositives}}$ (6)

Recall: The fraction of correctly predicted DDIs over all actual DDIs.

Formula: $\text{Recall}\frac{\text{TruePositives}}{\text{TruePositives+FalseNegatives}}$ (7)

Accuracy: The fraction of correct predictions and overall predictions.

Formula: $\text{Accuracy}\text{Correct Predictions Total Predictions}$ (8)

F1-score: The harmonic mean between precision and recall, providing an even measure of model performance.

Formula: $\text{F1} - \text{Score} = \frac{2\cdot\text{Precision}\cdot\text{Recal}}{\text{Precision+Recall}}$ (9)

These metrics are calculated based on the following definitions:

- True Positive (TP): Successfully identified DDIs.

- False Positive (FP): DDIs misidentified.

- True Negative (TN): Non-DDIs predicted correctly.

- False Negative (FN): Unable to predict DDIs.

Since precision and recall are likely to be inversely proportional, i.e., one may increase at the expense of the other, we used the F1 score to achieve a balanced metric. This assesses performance more rigorously, especially when working with imbalanced data.

This evaluation strategy ensures extensive testing of the predictive capability of the model and correcting for class imbalance issues.

A. Feature Importance Ranking:

In an attempt to enhance even further the trade-off between precision and recall, we utilized feature selection via the use of the Random Forest (RF) algorithm as well as Recursive Feature Elimination with Cross-Validation (RFECV). From the resultant features extracted from the dataset, we automatically identified and ranked the most informative features. From the original feature set, we retained the 500 most significant features regarding their influence on model performance. This action improves the ability of the model to accurately identify interaction types. The final data set is the selected features alongside important identifiers (e.g., drug pairsand cluster labels), ensuring that the most beneficial features are utilized for training and testing.

Model Performance:

Through this research, a drug-drug interaction (DDI) prediction model performance is enhanced for hypertension and diabetes medications utilizing advanced techniques to manage class imbalance. We applied SMOTEENN, a hybrid algorithm that applies oversampling to minority classes by removing noisy samples, which significantly improved the predictive accuracy of infrequent drug interactions. The results revealed that XGBoost outperformed other algorithms, including Random Forest (RF), Support Vector Machine (SVM), and Decision Tree (DT), as shown in TABLE 3. The application of SMOTEENN greatly improved performance by balancing class distribution, something that is of utmost importance while dealing with rare interactions between hypertension and diabetes drugs.

Algorithm	**Precision**	**Recall**	**F1-score**
Random Forest	0.94	0.92	0.93
XGBoost	0.98	0.97	0.97
Support Vector Machine	0.84	0.84	0.84
Decision Tree (DT)	0.85	0.85	0.85

Table 3: Algorithm Performance After Applying SMOTEENN

By 5-fold cross-validation, the XGBoost model consistently provided the highest precision, recall, and F1 score, establishing its exceptional performance for predicting intricate drug interactions.

Before incorporating SMOTEENN, the performance of the model on minority classes was poor, with recall values of less than 0.5 for certain types of interactions. Upon incorporating SMOTEENN, there was a significant improvement in identifying scarce interactions, leading to improved overall model accuracy.

Key improvements of SMOTEENN are:

1. **Enhanced Minority Class Representation:** By oversampling rare drug interaction pairs, the model learned more from minority classes.

2. **Noise Reduction:** Removing noisy and mislabeled samples improved recall and precision for all types of ininteractions.

3. **Consistency in Performance:** The model had balanced accuracy for both frequent and rare drug interactions, as shown in TABLE 4.

These findings support that the combination of XGBoost and SMOTEENN is the optimum way to maximize the prediction of drug interaction between hypertension and diabetes. The model generates a valid tool for the determination of possible adverse interactions, hence enhancing patient safety with reduced complications arising from medications.

Model Performance on Key Interaction Types:

Class	Precision	Recall	F1-Score	Support
0	0.94	0.90	0.92	316
1	0.98	0.96	0.97	461
2	0.96	0.95	0.96	368
3	0.98	0.88	0.93	312
4	0.96	0.95	0.96	602
5	0.98	0.98	0.98	661
6	0.97	0.99	0.98	763
7	0.97	0.99	0.98	799
8	0.98	0.99	0.98	836
9	0.98	0.99	0.98	749
10	1.00	1.00	1.00	863
11	1.00	1.00	1.00	835
12	0.99	1.00	0.99	841
Accuracy	-	-	0.98	8406
Macro Avg	0.98	0.97	0.97	8406
Weighted Avg	0.98	0.98	0.98	8406

Table 5: Model performance on key interaction types (XGBoost)

Training Accuracy	Test Accuracy
1.0000	0.9786

Table 5: Training And Test Accuracy of the Xgboost Model

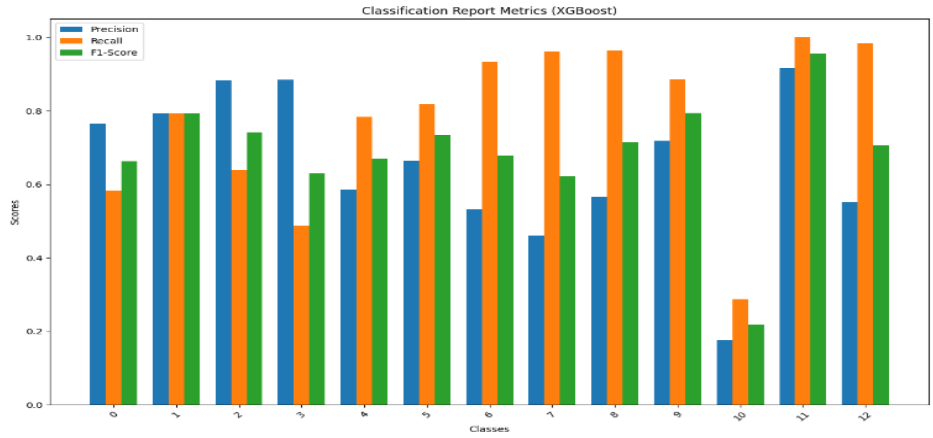

Figure **5**

This figure illustrates the performance of the model of XGBOOST model.

This figure illustrates the performance of the model of XGBOOST when predicting specific classes of drug-drug interactions that are clinically important. The performance metrics demonstrate the model's ability to correctly identify these interaction types, highlighting its effectiveness in distinguishing between high-risk and lower-risk combinations.

Class	**Precision**	**Recall**	**F1-Score**	**Support**
0	0.70	0.72	0.71	630
1	0.72	0.77	0.75	664
2	0.85	0.67	0.75	664
3	0.71	0.52	0.60	600
4	0.79	0.75	0.77	632
5	0.89	0.81	0.85	656
6	0.81	0.94	0.87	657
7	0.80	0.95	0.87	630
8	0.88	0.94	0.91	680
9	0.88	0.89	0.89	633
10	1.00	1.00	1.00	617
11	0.99	1.00	0.99	677
12	0.93	1.00	0.96	657

Metric	**Value**
Accuracy	0.84
Macro Avg	0.84
Weighted Avg	0.84

Table 6: Model Performance on Key Interaction Types (SVM)

Training Accuracy	Test Accuracy
0.8841	0.8440

Table 7: Training and test accuracy of the SVM model

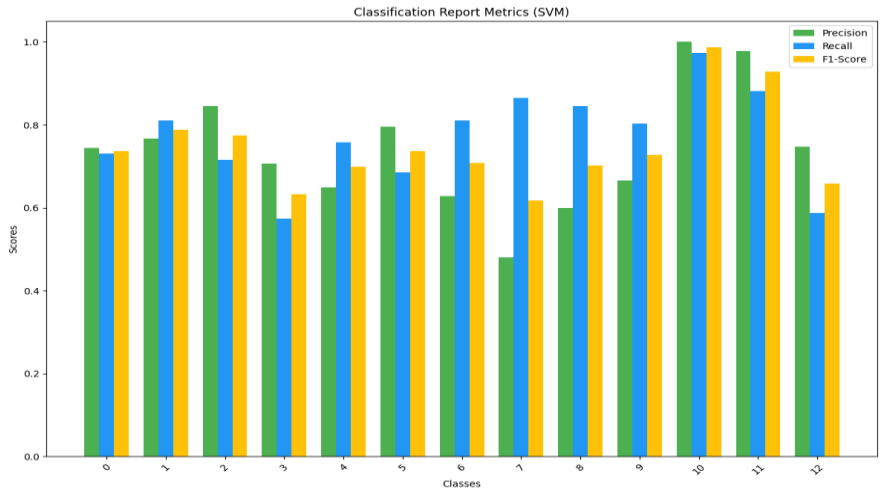

Figure 6

This figure illustrates the performance of the model of SVM.

This figure illustrates the performance of the model of SVM when predicting specific classes of drug-drug interactions that are clinically important. The performance metrics demonstrate the model's ability to correctly identify these interaction types, highlighting its effectiveness in distinguishing between high-risk and lower-risk combinations.

Class	Precision	Recall	F1-Score	Support
0	0.74	0.73	0.74	316
1	0.83	0.86	0.84	461
2	0.77	0.71	0.74	368
3	0.67	0.61	0.64	312
4	0.78	0.79	0.79	602
5	0.80	0.81	0.81	661
6	0.90	0.91	0.90	763
7	0.89	0.90	0.89	799
8	0.87	0.86	0.86	836
9	0.89	0.87	0.88	749
10	1.00	1.00	1.00	863
11	1.00	1.00	1.00	835
12	0.92	0.97	0.95	841

Metric	Value			
Accuracy	0.88			
Macro Avg	0.85			
Weighted Avg	0.88			

Table 8: Model Performance on Key Interaction Types (DT)

Training Accuracy	**Test Accuracy**
0.9487	0.8784

Table 9: Training And Test Accuracy of the TR Model

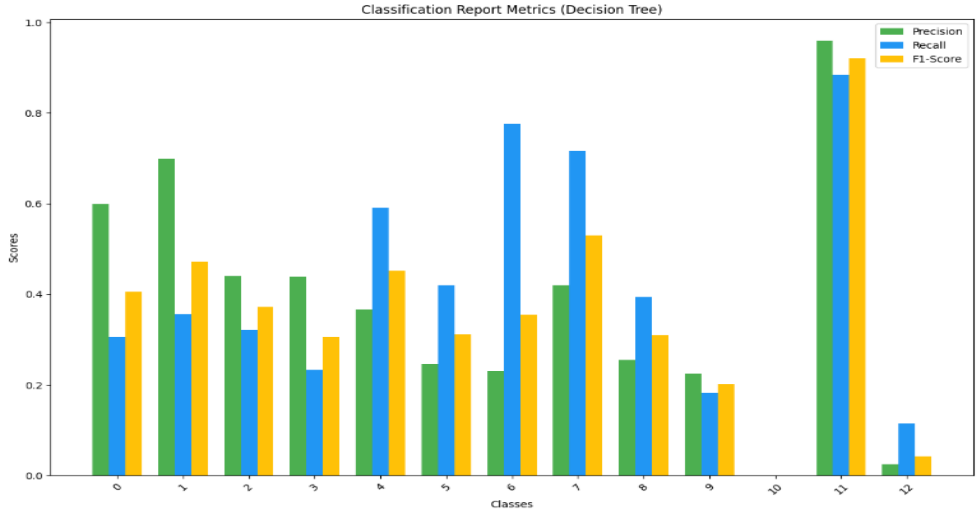

Figure 7. This Figure Illustrates the Performance of the Model of DT.

This figure illustrates the performance of the model of DT when predicting specific classes of drug-drug interactions that are clinically important. The performance metrics demonstrate the model's ability to correctly identify these interaction types, highlighting its effectiveness in distinguishing between high-risk and lower-risk combinations.

Class	**Precision**	**Recall**	**F1-Score**	**Support**
0	0.90	0.84	0.87	316
1	0.92	0.92	0.92	461
2	0.91	0.81	0.86	368
3	0.97	0.68	0.80	312
4	0.90	0.90	0.90	602
5	0.91	0.92	0.92	661
6	0.93	0.97	0.95	763
7	0.92	0.98	0.95	799

8	0.93	0.95	0.94	836
9	0.94	0.95	0.95	749
10	1.00	1.00	1.00	863
11	0.99	1.00	1.00	835
12	0.97	1.00	0.98	841
Accuracy			0.94	8406
Macro Avg	0.94	0.92	0.93	8406
Weighted Avg	0.94	0.94	0.94	8406

Table 10: Model Performance on Key Interaction Types (Random Forest)

Training Accuracy	**Test Accuracy**
0.9927	0.9424

Table 11. Training And Test Accuracy of the Random Forest Model

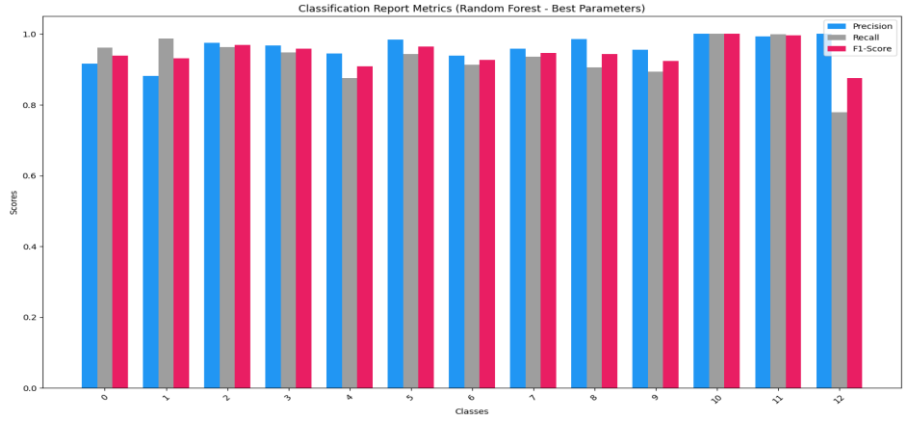

Figure 8. This Figure Illustrates the Performance of the Model of Random Forest.

This figure illustrates the performance of the model of Random Forest when predicting specific classes of drug-drug interactions that are clinically important. The performance metrics demonstrate the model's ability to correctly identify these interaction types, highlighting its effectiveness in distinguishing between high-risk and lower-risk combinations.

Model Performance Comparison:

For fully assessing the performance accuracy of the machine learning models for drug-drug interaction (DDI) prediction, we computed training and testing accuracy for XGBoost, SVM, Random Forest, and Decision Tree. It can be observed from Figure [X] that the maximum training accuracy (100%) was exhibited by the XGBoost model, indicating it possesses a class-leading ability to learn high-order relationships between the dataset. However, its test accuracy (97.86%) suggests that while the model generalizes well, there may be a small amount of overfitting. The Random Forest model also had high accuracy, with 99.27% on the training set and 94.24% on the test set, which suggests a strong balance between learning and generalization.

On the other hand, the SVM model had a large gap between training accuracy (88.41%) and test accuracy (84.40%), indicating poor generalization capacity and potential underfitting. The

Decision Tree model was 94.87% accurate in the training set but dropped to 87.84% in the test data, as shown in Figure 9, indicating an overfitting tendency due to the simplicity of the form.

In general, the Random Forest and the XGBoost models were the most effective algorithms for DDI prediction, of which the best accuracy was obtained with XGBoost at the expense of low overfitting risk. The findings emphasize the advantage of ensemble models in the handling of complex biochemical data and the improvement in predictive precision.

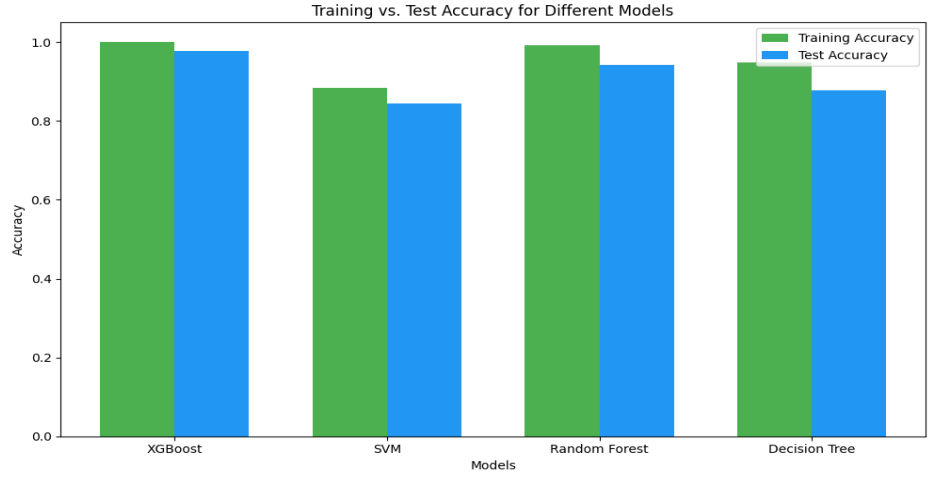

Figure 9. Comparison of Model Performance in Terms of Training and Testing Accuracy.

Comparison of model performance in terms of training and testing accuracy. The figure compares the training and test accuracy of four machine learning models, XGBoost, SVM, Random Forest, and Decision Tree, used for drug-drug interaction prediction. XGBoost achieved the highest training accuracy, indicating strong learning capacity, while its test accuracy reflects good generalization with slight overfitting. Random Forest also demonstrated strong performance with a good balance between training and testing accuracy.

The comparison plot simply illustrates the performance gap of the Random Forest, XGBoost, Support Vector Machine (SVM), and Decision Tree (DT) models based on three major evaluation metrics: Precision, Recall, and F1-score. Of note is the fact that the XGBoost model attained the highest scores in all the measures, which reflects its better ability to generalize and accurately classify information. The Random Forest model also demonstrated good performance, particularly after it was optimized in terms of hyperparameters and selection of the most relevant features. For comparison, SVM and Decision Tree models reported relatively lower scores, emphasizing the value added from using Random Forest for hyperparameter tuning and selection of features towards improving the model's accuracy and reliability. The visualization is efficient in showing the contribution of hyperparameter tuning to improving model performance according to varying evaluation metrics as shown in Figure 10.

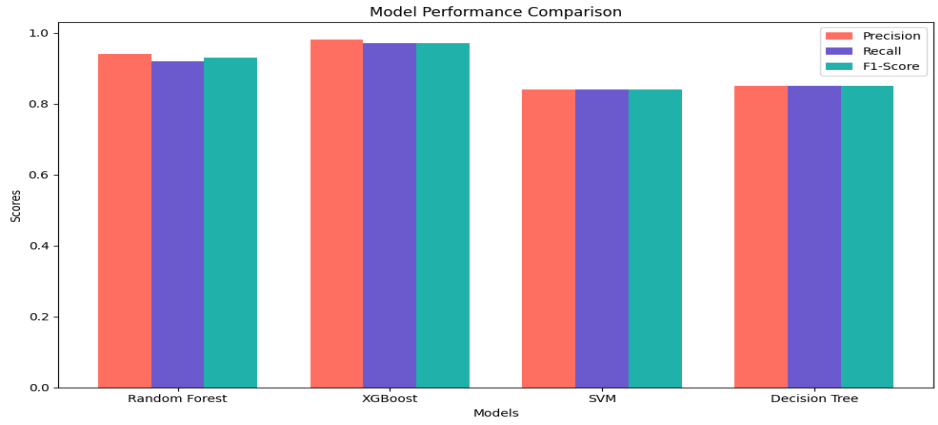

Figure **10**. Model Performance Comparison Based on Precision, Recall, And F1-Score.

Model performance comparison based on Precision, Recall, and F1-score. This figure illustrates the performance gap among XGBoost, Random Forest, SVM, and Decision Tree models across three key evaluation metrics. XGBoost consistently outperformed the other models, achieving the highest values across all metrics, followed by the optimized Random Forest model. The relatively lower scores of SVM and Decision Tree emphasize the effectiveness of advanced models and hyperparameter tuning in enhancing classification accuracy and reliability. Precise prediction of drug-drug interactions (DDIs) is very important in the treatment of diabetes and hypertension because medications for these diseases are commonly co-administered. In this research, it was proven that the integration of SMILES-based features and advanced machine learning algorithms offers a powerful framework for detecting possible interactions. Our approach used SMOTEENN to deal with class imbalance and XGBoost due to its high performance when dealing with big and complicated data. Another common issue for DDI prediction is class imbalance, where a minority of drug pairs are more common than others. If left uncorrected, the model would disregard minority but clinically relevant interactions. SMOTEENN largely mitigated this imbalance by oversampling minority classes through the elimination of noisy instances, achieving better precision and recall for both types of interactions. This tweaking was particularly beneficial in enhancing sparsely represented DDI classes in anti-hypertensive and diabetes drugs. The XGBoost model was more accurate compared to other models due to its ability to deal with complex relationships between drugs. Its ensemble learning-based approach reduces overfitting and provides a generalized prediction model, thus suitable for heterogeneous datasets. In applying our experiments, XGBoost was better compared to other machine learning models, especially when combined with SMOTEENN, providing a more balanced performance in both frequent and infrequent interaction classes. Our results are consistent with other research citing the necessity of incorporating chemical and pharmacological attributes for precise DDI prediction. SMILES representations enable easy extraction of chemical patterns, enabling the model to identify structural similarities and how they influence drug interactions. In contrast to conventional approaches based on chemical similarity measures such as the Tanimoto coefficient (Tc), our model utilizes machine learning to identify non-linear interactions among drug features, improving predictive power.

While promising, there are some limitations to be noted. Model performance is, of course, contingent on the dataset diversity and quality. While SMOTEENN effectively balances the dataset, interactions with sparse representation remain a challenge. Future improvements may involve adding other clinical datasets (e.g., real patient data) and molecular descriptors to provide a broader biochemical context. Besides, using deep learning techniques such as graph neural networks could further improve the model's sensitivity to identify nuanced interaction patterns. Lastly, our study highlights the practicality of combining SMOTEENN with XGBoost in the prediction of DDI among hypertension and diabetes medications. The approach not only improves model performance but also offers a scalable and effective platform for identifying potential drug interactions, ensuring safer drug handling and better patient outcomes.

Limited Set of Features:

Although the referenced model utilizes an acceptable set of chemical descriptors derived from SMILES representations, the model is limited in scope with its exclusive dependence on molecular and structural features. These are chemical formulae, physicochemical descriptors, biological activity, and structural interaction features. Although the features mentioned above are critical for computational modeling of drug molecules, they cannot well reflect the complexity of drug–drug interactions (DDIs) in real life.

DDIs are influenced by numerous biological, genetic, and clinical factors beyond molecular shape. The reliance on features derived from SMILES alone might result in a trivial model, constraining its accuracy in heterogeneous or clinically difficult scenarios. The current feature set misses some significant elements:

Absence of Pharmacogenomic Data:

Genetic variability is of crucial concern to an individual's ability to metabolize and respond to drugs. Omitting pharmacogenomic data from the model leaves it unable to address variability among individuals that may have potent influences on drug interactions. Incorporation of genomic data might provide a considerable input into increasing prediction accuracy and allow for the development of individualized DDI models.

Lack of Pharmacokinetic and Pharmacodynamic (PK/PD) Parameters:

Major PK/PD characteristics—such as absorption, distribution, metabolism, excretion (ADME), and the pharmacodynamic effects of drugs—are essential to understand how drugs behave in the body. These temporal and dynamic aspects are not captured in the current model, and thus it cannot simulate realistic interaction scenarios or forecast time-dependent side effects.

Binary Interaction Representation:

The model appears to forecast interactions as binary outcomes either an interaction occurs or does not. But drug interactions, when used in clinical practice, vary significantly in severity and cause. The simplification to binary can reduce the clinical utility of the model. Representing interaction severity or assigning interaction types would give more qualitative and informative outcomes.

Potential Data Bias and Limited Generalizability:

The training data from which the model is trained, being predominantly derived from DrugBank, can contain biases such as overabundance of common drugs and underabundance of rare or new interactions. Without independent validation on an external dataset, the strength and

generalizability of the model remain unknown.

Lack of Clinical Context:

The current feature set lacks critical clinical parameters such as comorbidities, age groups, and dosing intervals. These context-specific factors are involved in how interactions are presented in real-world scenarios. This limitation restricts the model's applicability, especially in complex patient cases such as patients with chronic diseases such as hypertension or diabetes.

Conclusion and Future Work:

In this study, we have demonstrated that the SMILES representation is an efficient and effective method for drug–drug interaction (DDI) prediction, particularly for drugs used to treat hypertension and diabetes. With the introduction of an Interaction Module, we were able to extract valuable features from a large-scale drug dataset, significantly accelerating the prediction process.

To fight class imbalance in the training and test data, we applied the SMOTEENN technique, and it led to improved model performance—particularly for underrepresented interaction classes. The technique achieved a good trade-off between precision and recall, which made the model more resilient to handling imbalanced data.

Among the models tested, the XGBoost algorithm was the most precise, owing to its efficiency in processing large, imbalanced datasets. The XGBoost with SMOTEENN was determined to be the optimal configuration, giving precise and reliable DDI predictions for drugs used in diabetes and hypertension treatment.

Despite these strengths, the model had somewhat poorer performance for low-frequency interaction types. Nevertheless, it was stable and consistent for most interaction categories and thus a valuable tool for the prediction of possible DDIs in clinical practice.

Future Directions

Based on the promising results of this study, future research will be focused on expanding the feature space and clinical usefulness of the model. While SMILES-based descriptors provided a solid foundation, drug interactions are influenced by a wider array of factors beyond molecular structure. In order to enhance predictive capability and clinical usefulness further, the following directions are proposed:

Integration of Pharmacogenomic Data: Genetic variation significantly affects drug metabolism and response. Incorporation of pharmacogenomic profiles will enable more personalized and accurate predictions of DDIs.

Inclusion of Pharmacokinetic/Pharmacodynamic (PK/PD) Features: Temporal and dynamic characteristics such as absorption, distribution, metabolism, and excretion (ADME), and drug effects on physiological systems, are crucial for the simulation of real-world drug activity.

Expansion of Clinical Context: Integration of the current biochemical features with real-world clinical data—e.g., comorbidities, age ranges, dosing schedules—will result in a more patient-specific and comprehensive model.

Use of External Validation Datasets: Use of diverse datasets outside of DrugBank will reduce potential biases and increase the generalizability of the model.

Advanced Representation of Interactions: Going beyond binary classification to include interaction severity and mechanistic pathways will increase the clinical usefulness and richness of the predictions.

Exploration of Deep Learning Techniques: Utilization of state-of-the-art methods such as Graph Neural Networks (GNNs) is extremely promising to represent complex molecular interactions with the possibility to enhance the model capacity in capturing subtle and high-order interaction patterns.

By filling these gaps, our goal is to create a next-generation DDI prediction framework that is not only highly accurate but also provides clinically relevant insights, with the ultimate goal of advancing safer and more effective drug therapy, especially for those patients suffering from chronic disorders like hypertension and diabetes.

Acknowledgement:

We would like to thank MEAPAL data Science @ CORELIA https://www.corelia.ai/ for their great collaboration and support.

References

Abacha, A. B., Chowdhury, M. F. M., Karanasiou, A., Mrabet, Y., Lavelli, A., & Zweigenbaum, P. J. J. o. b. i. (2015). Text mining for pharmacovigilance: Using machine learning for drug name recognition and drug–drug interaction extraction and classification. 58, 122-132.

Allahgholi, M., Rahmani, H., Javdani, D., Weiss, G., Módos, D. J. C. i. B., & Medicine. (2020). ADDI: Recommending alternatives for drug–drug interactions with negative health effects. 125, 103969.

Breiman, L. J. M. l. (1996). Bagging predictors. 24, 123-140.

Breiman, L. J. M. l. (2001). Random forests. 45, 5-32.

Breslow, R. A., Dong, C., White, A. J. A. C., & Research, E. (2015). Prevalence of alcohol-interactive prescription medication use among current drinkers: United States, 1999 to 2010. 39(2), 371-379.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining,

Cheng, F., & Zhao, Z. J. J. o. t. A. M. I. A. (2014). Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. 21(e2), e278-e286.

Deng, Y., Xu, X., Qiu, Y., Xia, J., Zhang, W., & Liu, S. J. B. (2020). A multimodal deep learning framework for predicting drug–drug interaction events. 36(15), 4316-4322.

DrugBank. (2025). Canadian Institutes of Health Research. Retrieved January 2, 2025 from https://go.drugbank.com

Farid, D. M., Zhang, L., Rahman, C. M., Hossain, M. A., & Strachan, R. J. E. s. w. a. (2014). Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks. 41(4), 1937-1946.

Fatehifar, M., & Karshenas, H. J. J. o. B. I. (2021). Drug-drug interaction extraction using a position and similarity fusion-based attention mechanism. 115, 103707.

Ferdousi, R., Safdari, R., & Omidi, Y. J. J. o. b. i. (2017). Computational prediction of drug-drug interactions based on drugs functional similarities. 70, 54-64.

Gottlieb, A., Stein, G. Y., Oron, Y., Ruppin, E., & Sharan, R. J. M. s. b. (2012). INDI: a computational framework for inferring drug interactions and their associated recommendations. 8(1), 592.

Ho, T. K. J. I. t. o. p. a., & intelligence, m. (1998). The random subspace method for constructing decision forests. 20(8), 832-844.

Hung, T. N. K., Le, N. Q. K., Le, N. H., Van Tuan, L., Nguyen, T. P., Thi, C., & Kang, J. H. J. M. i. (2022).

An ai-based prediction model for drug-drug interactions in osteoporosis and Paget's diseases from smiles. 41(6), 2100264.

Kuhn, M., Campillos, M., Letunic, I., Jensen, L. J., & Bork, P. J. M. s. b. (2010). A side effect resource to capture phenotypic effects of drugs. 6(1), 343.

Lee, G., Park, C., & Ahn, J. J. B. b. (2019). Novel deep learning model for more accurate prediction of drug-drug interaction effects. 20, 1-8.

Lounkine, E., Keiser, M. J., Whitebread, S., Mikhailov, D., Hamon, J., Jenkins, J. L., Lavan, P., Weber, E., Doak, A. K., & Côté, S. J. N. (2012). Large-scale prediction and testing of drug activity on side-effect targets. 486(7403), 361-367.

Lynch, T., & Price, A. J. A. f. p. (2007). The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects. 76(3), 391-396.

Park, C., Park, J., & Park, S. J. E. S. w. A. (2020). AGCN: Attention-based graph convolutional networks for drug-drug interaction extraction. 159, 113538.

Percha, B., & Altman, R. B. J. T. i. p. s. (2013). Informatics confronts drug–drug interactions. 34(3), 178-184.

Saverno, K. R., Hines, L. E., Warholak, T. L., Grizzle, A. J., Babits, L., Clark, C., Taylor, A. M., & Malone, D. C. J. J. o. t. A. M. I. A. (2011). Ability of pharmacy clinical decision-support software to alert users about clinically important drug—drug interactions. 18(1), 32-37.

Tatro, D. S. (2011). Drug interaction facts 2011: The authority on drug interactions. Wolters Kluwer Health/Facts & Comparisons.

Vapnik, V., & Izmailov, R. J. P. R. (2021). Reinforced SVM method and memorization mechanisms. 119, 108018.

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., & Sayeeda, Z. J. N. a. r. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. 46(D1), D1074-D1082.

Wu, H., Xing, Y., Ge, W., Liu, X., Zou, J., Zhou, C., & Liao, J. J. J. o. b. i. (2020). Drug-drug interaction extraction via hybrid neural networks on biomedical literature. 106, 103432.

You, J., McLeod, R. D., Hu, P. J. C. b., & chemistry. (2019). Predicting drug-target interaction network using deep learning model. 80, 90-101.

Zhu, Y., Li, L., Lu, H., Zhou, A., & Qin, X. J. J. o. b. i. (2020). Extracting drug-drug interactions from texts with BioBERT and multiple entity-aware attentions. 106, 103451.