SAKARYA UNIVERSITY JOURNAL OF COMPUTER AND INFORMATION SCIENCES
http://saucis.sakarya.edu.tr/
e-ISSN: 2636-8129
Publisher: Sakarya University
Vol. 8, No. 2, 245-259, 2025
DOI: 10.35377/saucis...1646658
SAUCIS
SAKARYA UNIVERSITY
RESEARCH ARTICLE

DeepInsulin-Net: A Deep Learning Model for Identifying Drug Interactions Leading to Specific Insulin-Related Adverse Events

Muhammed Ali PALA[1,2]

[1]Department of Electrical and Electronics Engineering, Faculty of Technology, Sakarya University of Applied Sciences, Sakarya, Türkiye; ORCID: 0000-0002-8153-7971
[2]Biomedical Technologies Application and Research Center (BIYOTAM), Sakarya University of Applied Sciences, Sakarya, Türkiye.

Corresponding author:

Muhammed Ali Pala,
Department of Electrical and Electronics Engineering,
Faculty of Technology,
Sakarya University of Applied Sciences,
Sakarya, Türkiye
pala@subu.edu.tr

Article History:
Received: 25.02.2025
Revised: 21.05.2025
Accepted: 11.06.2025
Published Online: 16.06.2025

ABSTRACT
Predicting clinical adverse effects resulting from drug-drug interactions is a critical research area for drug safety and patient health. Specifically, predicting adverse effects associated with insulin is crucial for clinical decision support systems and pharmacovigilance applications. This study proposes a deep learning-based model with high accuracy to predict adverse effects caused by drug interactions. In the literature, 17 different clinical side effects commonly associated with the hormone insulin have been identified. The properties of the drug molecules causing these interactions were calculated through MACCS, Morgan fingerprints and RDKit descriptors. These features are filtered by the variance thresholding method and optimized to improve classification performance. The model is built on a 1D CNN architecture that handles drug pairs as parallel inputs and a class weighting technique is used to eliminate class imbalance. Experimental results show that the model achieves 99.66% accuracy in training and 94.03% in validation, with training loss decreasing to 0.01 and validation loss stabilizing at 0.22. The ROC-AUC metric is above 0.99, indicating that the model can predict infrequent adverse events. The developed model provides a scalable, computationally efficient and highly reliable approach to predict the clinical consequences of drug interactions.

Keywords: Artificial intelligence, Personalized medicine, Drug-Drug interactions, Adverse drug events, Insulin, Deep learning

1. Introduction

In modern medical approaches, the use of drugs in treating and managing diseases is gradually increasing. In particular, chronic diseases are becoming widespread worldwide due to population aging and many other environmental and genetic influences [1]. This situation brings with it various challenges in the provision of health services and the treatment of diseases. Many alternative treatment methods and medicines are being developed to overcome these challenges [2]. In particular, polypharmacy, i.e., the regular use of more than one medication simultaneously, has become an essential step in today's treatment processes [3], [4]. Polypharmacy has the potential to treat many diseases in individuals simultaneously. However, while it has the potential to increase treatment success and manage comorbidities, it has also brought with it a significant increase in the side effects of drug-drug interactions that occur as a result of this multiple drug use [5], [6]. Drug-drug interactions occur when one drug alters another drug's pharmacokinetic or pharmacodynamic properties. These interactions can lead to a decrease in treatment efficacy, making it difficult to control the disease, as well as significantly increasing the risk of adverse drug reactions (ADRs) [7]. For example, one drug may decrease the efficacy of another by accelerating its metabolism, while another drug may increase its toxicity by slowing its metabolism. In clinical practice, this can manifest as unexpected treatment failures, increased side effects, the emergence of new symptoms or exacerbation of existing diseases [8]. The clinical burden of DDIs not only adversely affects treatment outcomes but also poses a serious threat to patient safety and a significant economic burden on healthcare systems [9]. Therefore, accurately predicting the clinically relevant consequences of drug interactions, mitigating risks, and developing effective management strategies have become critical global priorities to improve patient care quality, ensure patient safety, and reduce healthcare expenditures. In this context, there is a growing need for innovative approaches to detect drug interactions early and accurately, support clinical decision-

Cite as: M. A. Pala. (2025). DeepInsulin-Net: A deep learning model for identifying drug interactions leading to specific insulin-related adverse events. *Sakarya University Journal of Computer and Information Sciences*. vol. 8, no. 2, pp. 245-259, 2025 doi: 10.35377/saucis...1646658

making and improve patient outcomes.

Insulin is a vital hormone with pleiotropic effects in the human body. Produced by the beta cells of the pancreas, insulin primarily plays a central role in regulating glucose metabolism. It controls blood glucose levels by ensuring glucose is taken into cells and used for energy production [10], [11]. However, the effects of insulin are not limited to glucose metabolism. It also plays essential roles in many physiological processes, such as protein and fat metabolism, cell growth and differentiation, vascular function and even neuronal activity [12]–[14]. This wide range of physiological effects makes insulin important in maintaining the body's homeostatic balance [15]. This vital role of insulin and its wide range of actions are directly related to various clinical conditions resulting from its deficiency, resistance or interactions with other drugs [16]. These conditions are not limited to a generalized picture of insulin deficiency or insulin resistance but can lead to a wide range of specific and clinically identifiable diseases and disorders [17], [18]. These diseases and conditions range from hypoglycemia to hyperglycemia, from diabetic ketoacidosis to chronic complications such as diabetic nephropathy, diabetic neuropathy and diabetic retinopathy, from gestational diabetes to neonatal hypoglycemia [19], [20]. Even less well-known but essential clinical conditions, such as glucose intolerance, abnormal liver function tests (LFTs), polydipsia, polyuria and peripheral vascular disease, can result from insulin dysfunction or drug interactions [21]. The diversity and complexity of these clinical outcomes make it necessary to consider these conditions and diseases as insulin-related and carefully examine drug interactions in this area.

Each of these insulin-related diseases and conditions implies different levels of risk and treatment management challenges for patients. For example, hypoglycemia is an acute condition that can be serious, even fatal, especially in elderly and frail patients [22]. Chronic complications, such as diabetic nephropathy and retinopathy, are conditions that significantly reduce the quality of life in the long term and cause severe morbidity and mortality [23]. Conditions such as gestational diabetes and neonatal hypoglycemia are critical for both maternal and infant health and require specialized treatment approaches [24]. Therefore, accurately predicting and classifying the potential clinical consequences of drug interactions in the context of these insulin-related diseases and conditions is vital to personalize treatment strategies, reduce patient risk and support clinical decision-making [9]. The clinical diversity, importance, and potential impact of drug interactions on these conditions are the primary motivations for our study, which focuses on insulin-related drug interactions and their clinical consequences.

Timely and accurate detection of drug-drug interactions is critical to ensure patient safety and optimize treatment success [25], [26]. Traditional DDI prediction methods rely heavily on pharmacological studies, clinical observations, adverse event reports and expert knowledge. While these methods are valuable in identifying interactions between specific drug pairs, they are inadequate in the face of today's rapidly expanding drug knowledge and the increasing reality of polypharmacy [27]. Traditional methods often struggle to uncover complex interaction patterns that require large-scale data analysis and are limited in generating reliable predictions for new drug combinations. Furthermore, these methods usually assess the overall interaction risk and do not provide detailed information on the specific clinical consequences of the interaction [28]. This makes it difficult for clinicians to personalize treatment decisions and minimize patient risk.

Artificial intelligence, especially deep learning methods, has revolutionized drug discovery and development processes in the biomedical field in recent years [29]–[31]. Through multilayer artificial neural networks, deep learning can automatically extract meaningful features and patterns from large and complex data sets. This ability has made deep learning a powerful tool for solving complex problems like drug-drug interaction prediction. Deep learning models have shown promising results in predicting DDIs by integrating information from various data sources such as chemical structures of drugs, genetic data, biological pathway information, electronic health records and scientific literature [25], [32]. In the literature, deep learning-based DDI prediction models have achieved higher accuracy, sensitivity and specificity than traditional methods. These models can detect known interactions and predict novel and potential interactions. Deep learning also offers significant advantages in modeling multidrug interactions, predicting the effects of drug combinations and elucidating the mechanisms underlying interactions [33], [34]. However, most existing deep learning-based DDI prediction studies often focus on generic drug interactions and do not provide in-depth analyses of specific therapeutic areas or clinical outcomes. This limits the potential to provide more meaningful and actionable insights in clinical practice. This shortcoming reveals a significant research gap for better understanding and management of drug interactions in patients with diabetes and insulin therapy.

Among deep learning architectures, Convolutional Neural Networks (CNNs) are emerging as a notable alternative in predicting drug-drug interactions [35]. Although CNNs were first recognized for their superior performance in image processing, especially one-dimensional Convolutional Neural Networks (1D CNNs), they can be applied successfully to one-dimensional data encoding the molecular properties of drugs. 1D CNNs can automatically learn local patterns and hierarchical relationships in the data by performing a convolution operation through a convolutional filter acting on the input data. This capability significantly reduces the need for manual feature engineering and allows the model to delve deeper into complex interaction mechanisms that have not been previously identified. This offers a critical advantage for modeling the subtle and complex biochemical processes underlying drug interactions.

2. Proposed Method

This study presents a novel and advanced deep-learning approach that aims to classify potential insulin-related clinical side effects caused by drug-drug interactions with high accuracy and specificity. The flowchart of the proposed method is given in Figure 1. The technique, DeepInsulin-Net, examines 17 clinical side effects most commonly associated with insulin-related adverse events found in the literature. This multi-class classification approach allows the model to learn more complex interaction patterns and make more accurate predictions. The representation of drug molecules through SMILES (Simplified Molecular Input Line Entry Specification) using a feature set enriched with various fingerprints and identifiers allows the 1D CNN model to consider different chemical and structural properties. The parallel 1D CNN architecture of the model enables the features of both drugs to be processed independently and then combined to predict the probability of interaction. This approach reduces the computational cost and increases the scalability of the model. By combining a customized 1D CNN architecture for the multiclass classification problem, an enriched molecular feature set, and a focus on insulin-related drugs, the study presents a novel and technically powerful method in the field of DDI prediction.

The proposed method provides a multi-stage, integrated pipeline that combines the power of a 1D CNN architecture with a rich feature set representing drug molecular structures and properties. In the first stage of this pipeline, data on insulin and related drug pairs and their interactions were obtained from TWOSIDES, a comprehensive drug interaction database. This dataset was subjected to rigorous pre-processing, including identifying and cleaning conflicting labels. The interaction labels were converted into a numerical format according to the requirements of the model, and the dataset was divided into training and validation sets to evaluate the model's generalization ability. An extensive feature engineering process was applied to capture the complex properties of drugs at the molecular level and improve the model's classification performance. In this process, molecular fingerprints representing the chemical structures of drugs and various molecular descriptors were computed using the RDKit library. These features provide rich and multidimensional information about the chemical structures, physicochemical properties and potential interaction mechanisms of drugs. These features obtained from different sources were combined into a single feature vector by concatenation and scaled in the range [0, 1] to make them suitable for the model's input. To reduce the model's complexity and select the most informative features, feature selection was performed using the variance threshold method.

The model's architecture consists of 1D convolutional layers parallel to each drug pair, which learn the latent and hidden representations of the drugs by taking molecular feature vectors as input. These representations for the two drugs are combined through a concatenation layer and transferred to a standard dense layer. This layer is connected to a softmax output layer that estimates the interaction probability based on 17 different classes of clinical outcomes. The performance of the developed model is extensively evaluated on an independent validation dataset. The overall performance of the model and its ability to predict different clinical outcomes were analyzed using various performance metrics such as accuracy, AUC, precision, recall, and confusion matrix. The clinical outcomes the model predicts with higher accuracy, in which classes it struggles, and the possible sources of errors are analyzed in detail. As one of the first studies to focus on classifying the clinical outcomes of insulin-related drug interactions, this study makes a unique and valuable contribution to the DDI prediction literature. Integrating the developed model into clinical decision support systems has the potential to optimize medication management, reduce the risk of adverse events and improve patient safety in patients with diabetes and insulin therapy.

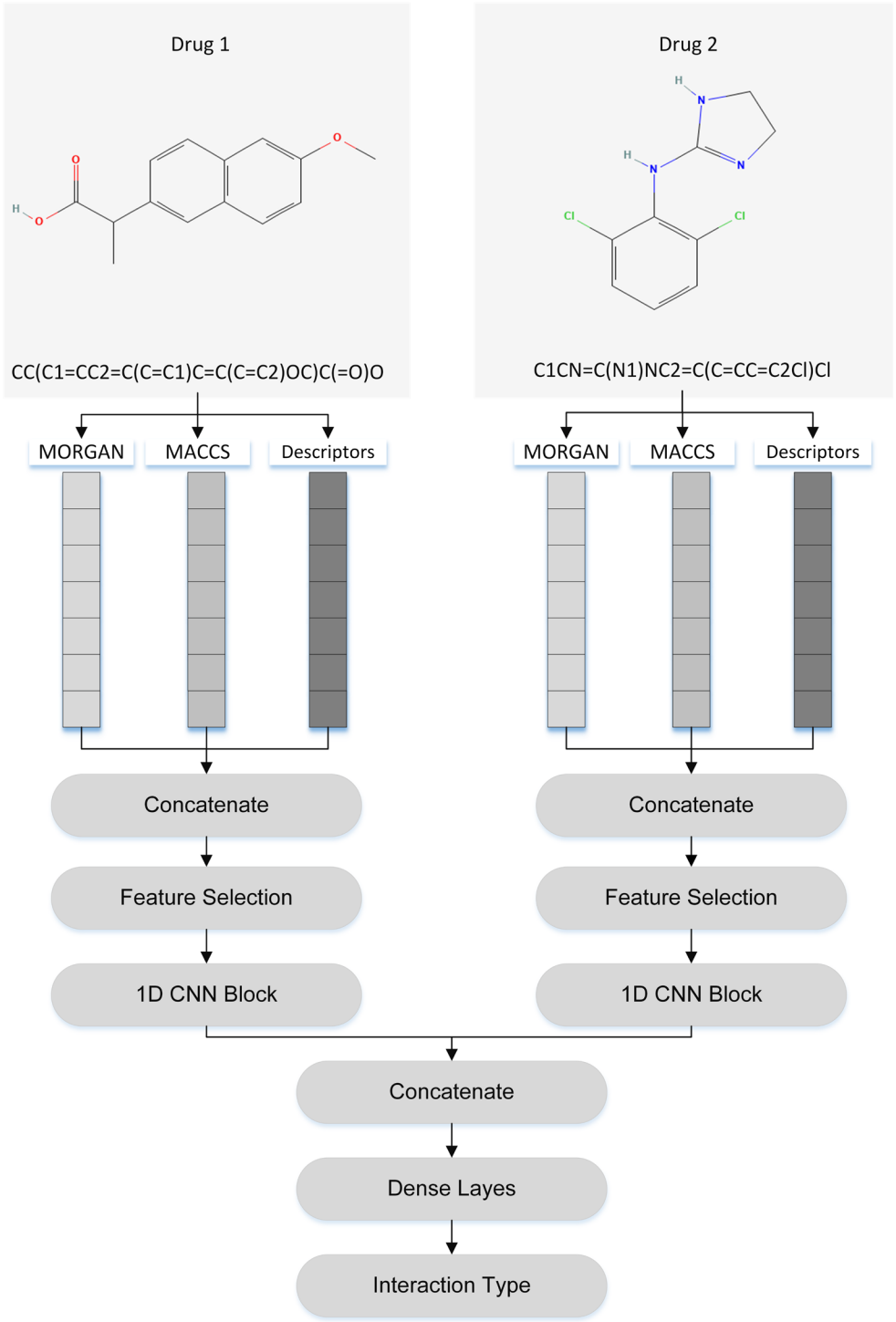

Figure 1. Flowchart of the DeepInsulin-Net.

3. Method

3.1 Data set and pre-processing

This study used the TWOSIDES dataset, a comprehensive and reliable drug interaction database derived from clinical data, to classify the clinical outcomes of insulin-related drug-drug interactions [36]. The TWOSIDES database contains 868,221 significant associations and 1,301 adverse events for 59,220 different drug pairs. These adverse events include conditions that cannot be attributed to any single drug. Furthermore, the database contains 3,782,910 significant associations of drug pairs, and the adverse effect scores of these interactions were higher than those of the individual drugs involved. In our study, we mainly focused on developing models that predict adverse drug reactions resulting from synergistic drug-drug interactions. The fact that TWOSIDES covers a wide range of drugs and drug interactions ensures that interactions of insulin

and related drugs are adequately represented in the database, providing a rich and reliable data source suitable for the study. The TWOSIDE dataset includes the SMILES representation of each drug pair to identify drug interactions and the type of drug interaction as label information.

A series of systematic steps were followed to select insulin-related side effects from the TWOSIDES database. First, insulin-related side effects and interactions were carefully identified through a literature search. Then, from the TWOSIDES database, all drug pairs containing these identified drugs and adverse events resulting from their interactions were eliminated from the entire dataset using text-mining steps. These adverse events were classified into 17 different clinical outcome classes, which were the focus of the study. These classes comprehensively included diabetic complications (diabetic nephropathy, diabetic retinopathy, diabetic neuropathy), disorders of glycemic control (hypoglycemia, hyperglycemia), gestational diabetes, abnormal liver function tests and other related conditions. This raw data was transformed into a structured dataset, with each row indicating a drug pair and the clinical outcome resulting from their interaction. The DDI interactions analyzed in the study are given in Table 1. In total, 93626 interaction pairs were used for the proposed method. The dataset was randomly split into training and evaluation in a 7:3 ratio.

Table 1. Number of insulin-related data in the DDI dataset.

Type	**Number of Interaction**
Hyperglycaemia	18779
Diabetes	12718
Hypoglycaemia	12443
Abnormal LFTs	10960
Polydipsia	5934
Diabetic neuropathy	5168
Peripheral vascular disease	5165
Diabetic acidosis	3523
Polyuria	3366
Insulin-dependent diabetes mellitus	3335
Diabetic Retinopathy	2982
Glucosuria	2894
Glucose intolerance	2107
Ketoacidosis	1362
Diabetic Nephropathy	1264
Hypoglycaemia neonatal	1206
Gestational diabetes	420

Furthermore, to ensure that the performance and reliability of the model are maximized, a comprehensive pre-processing process was applied to the resulting dataset. This process primarily involved identifying and cleaning records that reported multiple and different clinical outcomes for the same drug pair, i.e., conflicting labels. Removing these conflicting records from the dataset improved the consistency and learning ability of the model. The main reason for eliminating conflicting data is that the same drug pair can lead to multiple and different clinical outcomes, complicating the model's learning process and leading to misleading results. Therefore, SMILES strings, in this case, were removed from the dataset. Figure 2 shows the data distribution plot for 17 different insulin-related drug interactions. As can be seen, the dataset has an unbalanced class distribution. The proposed method uses a class weighting strategy to address the unbalanced class distribution. In this method, weights are assigned to each class based on the class distribution in the training data. In this way, it aims to take more imbalanced classes into account by the model and improve its learning process on imbalanced datasets. The obtained weights are integrated into the model's loss function to minimize the performance problems caused by class imbalance.

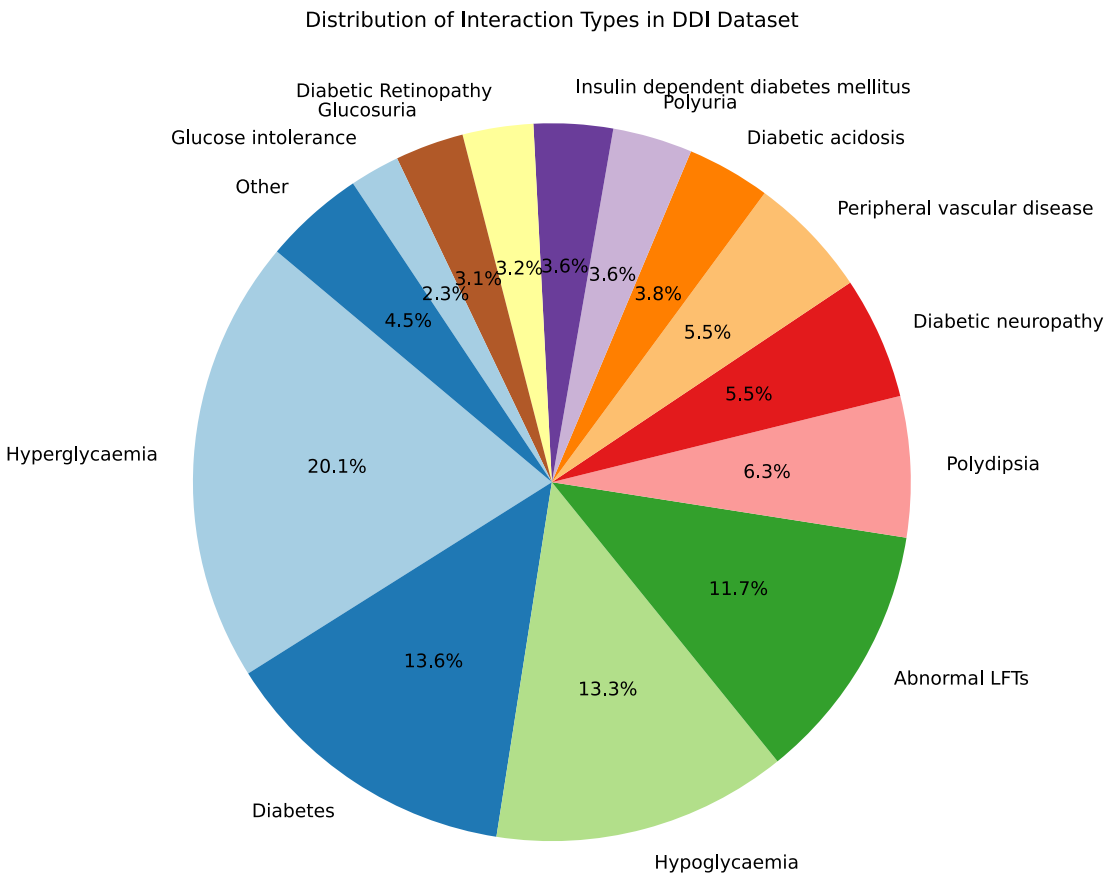

Figure 2. Data distribution

3.2 Fingerprints and Descriptors

Accurate, reliable and clinically meaningful prediction of drug-drug interactions is, without a doubt, critically dependent on representing the molecular structures of drugs and the properties derived from these structures in a way that is appropriate, informative and serves the learning capacity of the model. The transformation of drugs in molecular form into numerical vectors and abstract mathematical objects allows machine learning algorithms to learn the subtle but deeply complex relationships between drugs, make generalizations from these relationships, and ultimately predict interactions for unknown drug pairs. This study used different representation methods, namely two different molecular fingerprints and molecular descriptors, to represent the molecular properties of drugs in a way that encompasses both structural and physicochemical dimensions, i.e., by providing multidimensional and complementary information, thus maximizing the model's classification performance. These approaches offer vital information about the chemical structures, three-dimensional conformations, physicochemical properties and potential interaction mechanisms of the drugs, which are crucial for the model to classify DDIs successfully. MACCS and Morgan fingerprints of drug molecules, as well as RdKit descriptors, were used in this study. SMILES representations of the molecules were used to calculate the fingerprints and identifiers.

MACCS fingerprints are a well-established type of fingerprinting developed by MDL Information Systems, including 166 predefined structural keys [37]. These structural keys are carefully selected structural motifs representing functional groups, ring systems, and specific atomic numbers commonly found in drug molecules. The MACCS fingerprint of a molecule consists of a 166-bit array that indicates whether these 166 keys are present in the molecule. MACCS fingerprints have the advantages of covering an ample chemical space, being relatively interpretable and straightforward, and being computationally efficient. Morgan fingerprints, also known as circular fingerprints, are a type of fingerprint that takes into account the atom-centered peripheral properties of the molecule [38]. Starting from each atom, Morgan fingerprints form circular substructures that contain all neighboring atoms and bonds within a given radius. These substructures are converted into unique integer values using a hash function, and these integers are used to set the corresponding bits in the fingerprint vector to 1. The most crucial feature of Morgan fingerprints is their ability to capture in detail the local structural features of the molecule and the neighborhood and bonding patterns of atoms and, thus, more precisely, identify subtle structural similarities and differences between different molecules. In this study, a rich and detailed representation of the molecular structure of drugs was obtained using Morgan fingerprints with a length of 2048 bits and a radius of 2. Molecular descriptors

are another critical and complementary category of properties that express drug molecules' physicochemical, topological, electronic and geometrical properties in numerical values. Descriptors quantitatively characterize a wide range of properties of the molecule, such as size, shape, surface area, polarity, hydrophobicity, hydrogen bonding capacity, dipole moment, ionization potential and electronic charge distribution, which directly affect the pharmacokinetic and pharmacodynamic properties of the drug. In this study, using the open-source and widely used RDKit chemoinformatics library, four basic molecular descriptors, namely molecular weight (MolWt), LogP (octanol-water partition coefficient), number of hydrogen bond donors (NumHDonors) and number of hydrogen bond acceptors (NumHAcceptors), were calculated, which are frequently used in the literature and play an essential role in predicting the biological behavior of drugs [39].

These molecular fingerprints and descriptors calculated for each drug molecule were combined into a single, long feature vector by concatenation. The concatenated feature matrices resulting from this process, which are performed separately for each drug interaction pair, are applied as input to the CNN model. This combined feature vector represents the structural and physicochemical properties of drugs together and comprehensively, providing a rich and informative input that allows the model to classify drug-drug interactions more accurately, precisely and reliably.

3.3 Feature Selection

Feature selection is essential in data analysis and modeling processes because it can affect the model's learning capacity as the data's size and complexity increase. In deep learning algorithms, too many features can increase the training time of the model, cause overfitting and limit the model's generalization ability [40]. For this reason, only essential and meaningful features should be included in the model. Especially in high-dimensional datasets, some features may lead to performance loss instead of improving the model's accuracy. Therefore, feature selection is critical to ensure the model runs efficiently and accurately. The variance threshold feature selection method is applied to the combined feature matrix created with fingerprints and descriptors [41]. This way, the feature matrices of high-dimensional drug molecules are expressed in a more meaningful matrix format. Therefore, this reduced the computational burden of the model and improved the generalization performance of the model.

The Variance Threshold method calculates the variance of each attribute and removes features with variance below a specific threshold value. Variance is a statistical measure of how variable the values of an attribute are and is calculated as follows:

$$\text{Variance}(X_j) = \frac{1}{n-1}\sum_{i=1}^{n}(X_{ij} - \mu_j)^2 \qquad (1)$$

Where X_j represents the values of the feature across all samples. μ_j is the mean of the $j-th$ feature and n is the number of the samples. If a feature's variance is lower than a specified threshold (θ), that feature is removed from the dataset. The variance thresholding method usually aims to extract features with low variance in the dataset, i.e., features with similar values in all instances. Such features do not benefit the model or contribute to the learning process.

3.4 CNN Model

This study developed a novel and advanced deep learning model based on a 1D Convolutional Neural Network architecture to classify the clinical outcomes of insulin-related drug-drug interactions with high accuracy and specificity. CNNs first proved themselves in image processing, especially in object recognition, image classification and face recognition. However, the basic principles of CNNs, particularly the advantages of convolution, have made them successfully applicable to various fields such as natural language processing, text classification, sentiment analysis, time series prediction, anomaly detection, gene sequence analysis and protein structure prediction.

1D CNNs are a subtype of CNNs specialized for application to one-dimensional data. 1D data is data with a sequential structure, such as texts, audio signals, time series, gene sequences and, as in this study, molecular fingerprints and identifiers. 1D CNNs perform a convolution on such data using a sliding filter. This filter is a small matrix of weights that moves step by step along with the input data with a given stride size. At each step, an element-wise multiplication is performed between the input region covered by the filter and the filter weights, and the sum of the resulting products forms the corresponding element of a feature map [42]. This process enables the automatic detection of local patterns in the input data. The architecture of 1D CNNs typically consists of an input layer, one or more convolutional layers, activation functions, pooling layers, fully connected layers and an output layer. The input layer receives the raw data that the model will process. In this study, the input layer receives the combined feature vectors obtained during the feature extraction stage of drug feature engineering, which is dimensionally reduced by feature selection. Convolutional layers are the basic building blocks of 1D CNNs and produce feature maps by performing a convolution on the input data. Each convolutional layer uses different filters, each learning to capture different local patterns in the input data. Nonlinearity is introduced into the model by applying activation functions to the outputs of the convolutional layers. Activation functions enable the model to learn nonlinear relationships and solve more complex problems. Dense layers are usually located at the end of the CNN architecture and connect all neurons from the previous layers. These layers allow the local features learned by the convolutional layers to be combined to form higher-level and abstract representations. The output layer is where the model produces its final predictions. In this study, the output

layer computes the probabilities of 17 different clinical outcomes for each drug pair using the softmax activation function. The architecture of the 1D CNN model developed for classifying insulin-related DDI pairs is given in Figure 3.

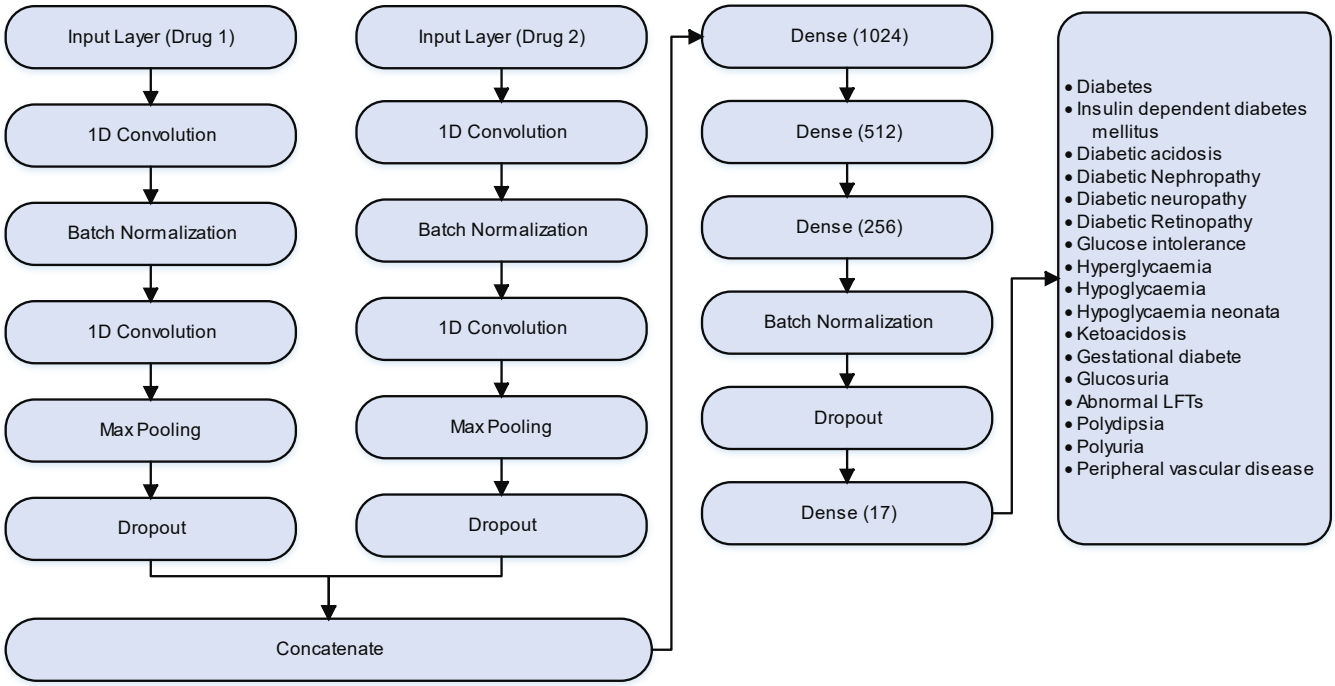

Figure 3. DeepInsulin-Net CNN model architecture.

The model proposed in this paper has a dual-input structure. It uses a parallel processing architecture to process two different drug feature matrices (Drug1 and Drug2) separately and then combine them to obtain a feature representation. In the model's design, each drug's data features are processed independently and combined to get the final classification result. Such a parallel structure allows for more efficient drug interaction modeling while ensuring that each drug's data features are considered separately. This approach offers significant advantages, especially in solving complex biological problems like drug interactions. The model independently processes feature vectors from two drugs to create high-level representations and combines these representations to perform the final classification. The parallel input structure processes the molecular features of both drugs through separate convolutional blocks, aiming to obtain independent but meaningful representations of each drug in a common feature space. This architecture offers higher generalization capacity when analyzing the potential for different drugs to interact with each other.

In the proposed model, there is an independent 1D CNN block for each drug. First, each input is processed with a convolutional layer. In the first convolutional block, feature extraction is performed with a 3D kernel using 128 filters. This is followed by a batch normalization process to provide more balanced learning in the model's training process. Then, the convolution layer is used again. After each convolution block, a dropout layer of 0.1 is added to avoid overfitting the model. After the first convolution block, a second convolutional layer extracts higher-level features with 64 filters. Similarly, this layer undergoes batch normalization, pooling, and dropout to complete the separate processing of each input. In the final stage, after the convolutional processing of each input is completed, it is converted into vector form through the smoothing layer and made ready for processing with dense layers.

In the fusion stage of the model, the feature vectors obtained from both inputs are combined using the concatenate layer. The concatenated vector was processed using the ReLU activation function in dense layers with 1024, 512 and 256 neurons, respectively. Batch normalization was applied to the output of this layer. In the output layer of the model, a dense layer with a softmax activation function was used to consider the multiclass classification problem. The model was compiled with the Adam optimization algorithm (learning rate: 0.0005), and categorical cross-entropy was used as the loss function. During the training process, an early stopping mechanism was added to monitor the model's performance and to activate the early stopping mechanism, which terminates training if the validation loss does not improve for 100 epochs. In the training process, a class weighting method was used to make the model learn in a more balanced way. If the class distribution is unbalanced, class weights are calculated from the training data and integrated into the model's loss function to increase the model's sensitivity to rare classes. The training process was completed after 50 epochs as it met the desired criteria. The hyperparameters of the designed model are given in Table 2.

Table 2. Hyperparameters of the proposed model

Hyperparameters	**Value**
Number of Filters	128, 64
Kernel Size	3
Pooling Size	2
Dropout Rate	0.1
Optimization Algorithm	Adam
Learning Rate	0.0005
Batch Size	64
Epoch	50
Activation Function	ReLU, Softmax
Loss Function	Categorical Crossentropy

The performance evaluation metrics of the developed model are crucial for assessing its overall accuracy and the effectiveness of the learning process. This is particularly important when working with datasets that exhibit an unbalanced distribution, as it necessitates using multiple metrics during the training and evaluation phases. Consequently, various performance metrics were employed to gauge the comprehensive success of the model. In this context, we calculated the ROC-AUC (Receiver Operating Characteristic - Area Under Curve), recall, accuracy, and log loss.

4. Experimental Results

The model's performance is analyzed with various metrics, and the effects of different data representation techniques on the performance are examined. In addition, statistical analysis of the prediction results is performed to understand the decision mechanism of the model better. The multi-class dataset labels prepared in the study were converted into binary strings using the one-hot encoding method. The labels were transformed with this method and used as the target variable of the model. A structure was created to compare the predicted probabilities with the actual labels. This approach also helped reduce potential problems, such as unbalanced class distribution in the dataset and ensured stable learning during the model's training process. The experiments used a NVIDIA GeForce GTX 3070 GPU with Intel Core i7-11700H CPU @ 4.90 GHz and 32 GB RAM.

ROC-AUC is a critical metric that measures how well the model can distinguish between positive and negative classes at different thresholds. An AUC value close to 1 indicates that the model is highly discriminative, while an AUC value close to 0.5 indicates that the model performs at the level of random guessing. In the evaluation step, attention was paid to including only samples with at least two different classes in the analysis. The AUC metric is a vital indicator for understanding the overall generalization capacity of the model trained and evaluated with unbalanced data, and it reveals the level of accuracy of the model and the stability of the decision function. In addition, a log loss metric was used to determine the model's error rate. The categorical cross-entropy loss function is adopted in the model's training, and the log loss value provided by this function is used to analyze how well the probabilities generated by the model match the actual labels. The log loss metric offers a detailed insight into how reliable the estimated probability distribution is, showing the extent to which the model manages uncertainty. The early stopping method was applied to prevent the model from being overfit. This mechanism stops training and restores the best weights if there is no improvement in the verification loss for a specific period. In our study, the validation loss metric was monitored, and if there was no improvement in 100 epochs, the model training was terminated early. The trained model weights were recorded at each epoch to make this process more efficient. In this way, the weights with the lowest validation loss obtained at the end of the training process were reloaded, and the model returned to its best performance. The early stopping mechanism prevented the model from unnecessarily over-training, increasing the validation error and overall generalization capacity. The model trained with this strategy allowed the training and evaluation process to be completed in 50 steps.

As a result, the data processing techniques, metrics, and early stopping mechanisms applied to improve the model's performance and make the training process more efficient have ensured that the experimental results are reliable. The overall success rate of the model was examined in detail within the framework of the specified metrics, and the optimization processes were improved with the validation data at each stage of the performance improvements. Thus, it is proved that the proposed model works effectively in classification tasks and achieves high accuracy rates. The performance metrics of the proposed model during training and validation are given in Figure 4.

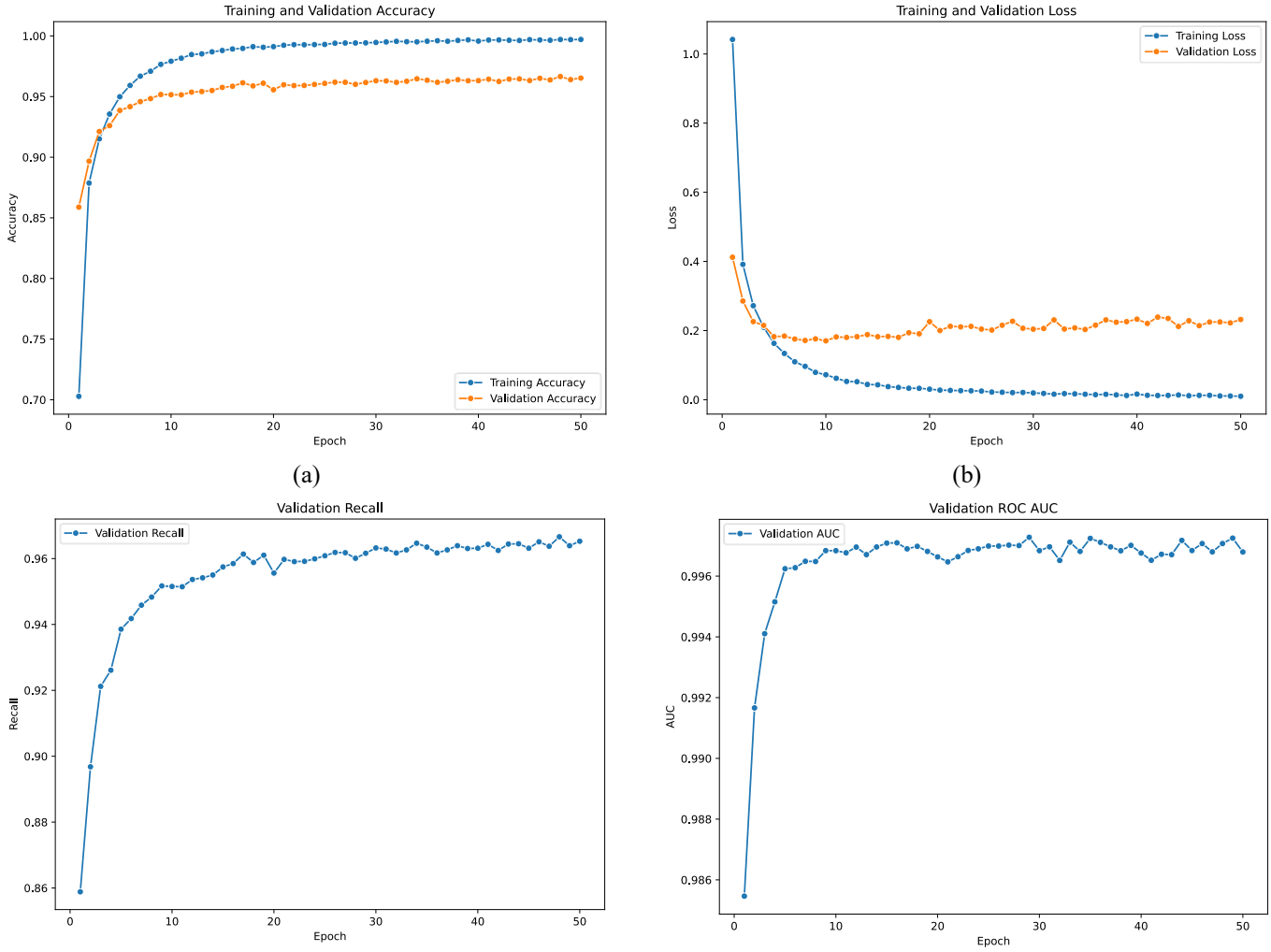

Figure 4. Performance metrics of the model were measured during training and evaluation

The performance of the developed 1D CNN model was comprehensively evaluated on the training and validation datasets through various metrics calculated for each epoch. The model's training runs for 50 epochs; the metrics obtained are presented in detail in Figure 3. The variation of the model's training and validation losses over epochs is visualized in Figure 3-b, while the variation of training and validation accuracies over epochs is visualized in Figure 3-a. The training loss of the model started from 1.04 in the first epoch, decreased steadily as the training process progressed and decreased to 0.01 in the last epoch. This shows that the model has successfully learned the patterns in the training dataset. Similarly, the training accuracy increased from 70.3% in the first epoch to 99.66% in the last. This high accuracy value shows that the model fits the training dataset well.

However, the model's performance on the validation dataset is the most important for assessing its generalization ability. Although the validation loss decreases in parallel with the training loss, it does not reach values as low as the loss. While the validation loss was 0.41 in the first epoch, it reached its lowest value at epoch 10 (0.17) and stabilized around 0.22 with slight fluctuations. The validation accuracy, however, increased from 85.9% in the first epoch to 96.1% in the 18th epoch and then stabilized around 96%. These results show that the model learned the patterns in the training dataset and successfully generalized them to the validation dataset, which it had not seen before. It is clear that the early stopping mechanism does not stop training before 50 epochs due to the halting of the improvement in the validation loss, and the model is trained to its optimum capacity. We also examined the AUC and recall values on the validation dataset to analyze the model's classification performance further. The validation AUC value is above 0.99 for all epochs, indicating that the model can successfully distinguish between positive and negative classes. The validation sensitivity values are also generally above 95%, indicating that the model's false-positive and false-negative predictions are very low. These results show that the model works with high precision, meaning that it both makes accurate predictions and does not miss actual positive samples. The confusion matrix, which shows a detailed overview of the model's correct and incorrect predictions for each clinical outcome class, is presented in Figure 5. The confusion matrix analysis allows for identifying which classes the model predicts better, which classes it struggles with, and which classes it confuses with each other. This analysis enables a better understanding of the strengths and weaknesses of the model.

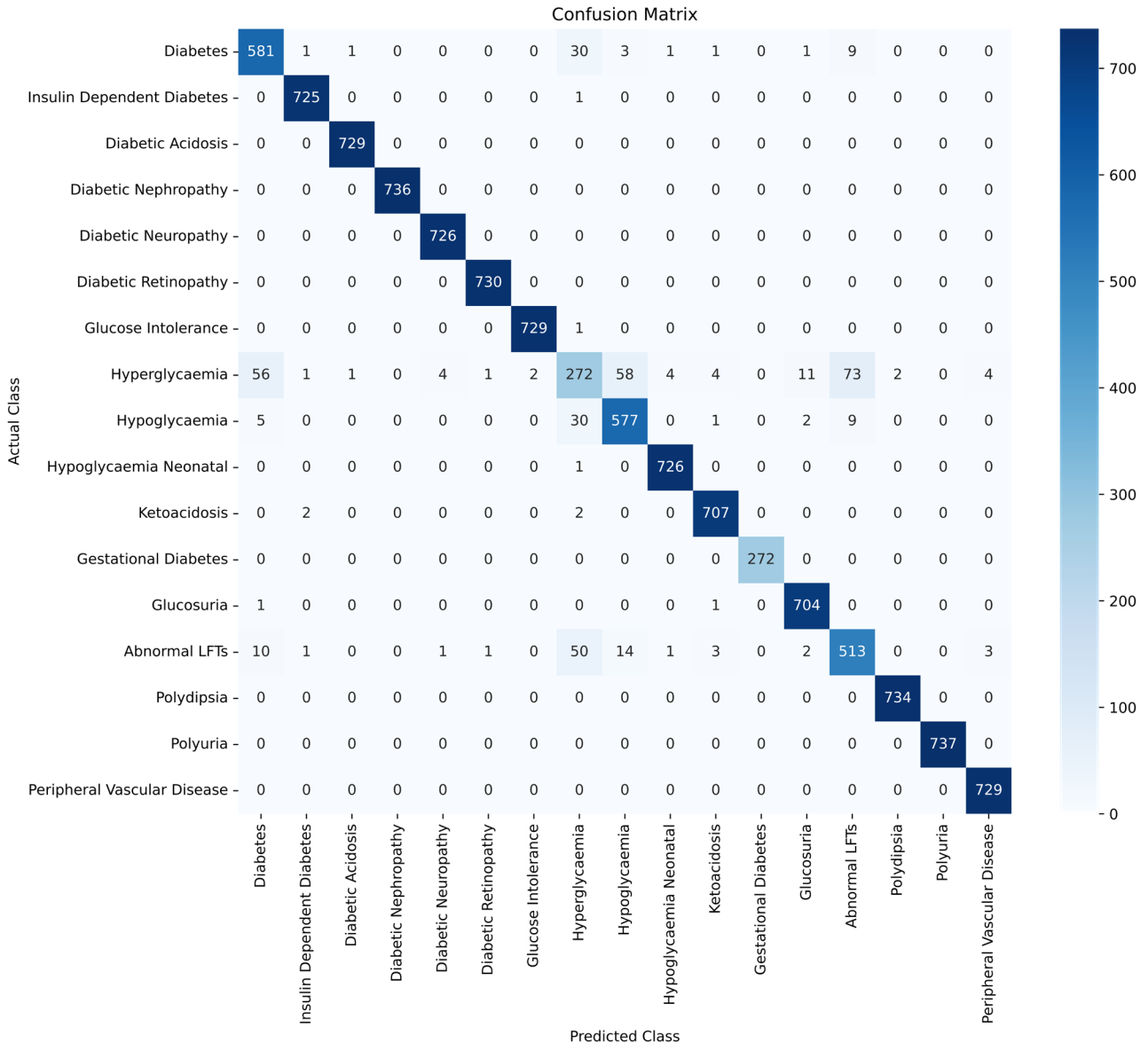

Figure 5. A confusion matrix of the proposed method was obtained during the validation.

The confusion matrix is widely used to evaluate the model's performance in multiclass classification problems. This matrix shows the relationship between the actual classes and the classes predicted by the model. The matrix rows represent the exact classes, and the columns represent the predicted classes. The cells on the diagonal represent the number of instances correctly classified by the model, while the cells off the diagonal represent the number of misclassified cases. As can be seen in Figure 5, the developed model has shown a highly successful classification performance in many different classes. The results show that the developed 1D CNN model can successfully classify the clinical outcomes of adverse effects of insulin-related drugs with high accuracy, sensitivity, and AUC values.

5. Ablation Study and Comparative Performance Evaluation

Several ablation analyses were conducted to evaluate the performance of the DeepInsulin-Net model and to measure the impact of pre-processing steps on classification performance. The performance of the developed model was systematically evaluated to assess the impact of data preprocessing strategies, such as feature selection and class weighting, on the predictive capabilities of various machine learning and deep learning models. The experimental evaluations were performed on a diabetes-related subset of the dataset containing clinically relevant interactions, and the data was stratified and split 7:3 for each scenario as training and validation, maintaining the class distribution. Once the molecular fingerprints and RDKit-based chemical identifiers of the drugs were calculated, four basic preprocessing scenarios were set up: an approach where feature selection and class weighting were used first; a case where only class weighting was applied, and all features were used; a scenario where only feature selection was performed, and class weighting was neglected; and an ablation analysis where no preprocessing steps were applied, and raw features were used. In addition to the proposed CNN model, these cases were

tested using Multilayer Perceptron (MLP), Support Vector Machines (SVM), and Extreme Gradient Boosting (XGBoost) algorithms. Model performances were measured for training and evaluation steps using accuracy, AUC, Cohen's Kappa and Matthews Correlation Coefficient (MCC). The results obtained from the ablation studies are given in Table 3.

Table 3. Performance of ablation analysis and other classification algorithms.

		Train				Validation			
Pre-Processing	Model	Accuracy	AUC	Kappa	MCC	Accuracy	AUC	Kappa	MCC
Feature Selection + Class Weights	CNN	**0,9966**	**0,9990**	**0,9964**	**0,9964**	**0,9403**	**0,9921**	**0,9259**	**0,9259**
	MLP	0,8319	0,8492	0,7702	0,8202	0,8080	0,7922	0,7423	0,7025
	SVM	0,8052	0,8479	0,7417	0,7917	0,7370	0,7387	0,5606	0,5233
	XGBoost	0,8288	0,8490	0,7669	0,8169	0,9111	0,7926	0,7456	0,7059
Class Weights	CNN	0,8136	0,8485	0,7507	0,8007	0,7470	0,7419	0,5712	0,5336
	MLP	0,8203	0,8487	0,7578	0,8078	0,7401	0,7410	0,5638	0,5260
	SVM	0,6733	0,8246	0,6016	0,6516	0,7924	0,7664	0,6195	0,5806
	XGBoost	0,8324	0,8492	0,7707	0,8207	0,9137	0,7929	0,7483	0,7087
Feature Selection	CNN	0,8436	0,8497	0,7826	0,8326	0,9168	0,7943	0,7516	0,7117
	MLP	0,6689	0,8233	0,5969	0,6469	0,7898	0,7654	0,6166	0,5781
	SVM	0,5323	0,7824	0,4518	0,5018	0,6519	0,7205	0,4701	0,4325
	XGBoost	0,6350	0,8140	0,5609	0,6109	0,7568	0,7543	0,5816	0,5439
Non-Pre-processing	CNN	0,6330	0,8122	0,5588	0,6088	0,7557	0,7548	0,5804	0,5425
	MLP	0,5323	0,7824	0,6518	0,6541	0,6519	0,7205	0,6301	0,6325
	SVM	0,6330	0,8122	0,7588	0,7607	0,7557	0,7548	0,7404	0,7425
	XGBoost	0,6689	0,8233	0,7969	0,7982	0,7898	0,7654	0,7766	0,7781

Analyzing the ablation results, it is clear that data preprocessing steps significantly impact classification performance. In the case of no pre-processing, they generally exhibit the weakest validation metrics, demonstrating the challenges posed by the class imbalance and high feature size in the raw data. In this baseline scenario, the XGBoost and SVM models performed slightly better than CNN and MLP, especially on the Kappa and MCC metrics, but their overall performance was limited. This shows that even without pre-processing, some traditional models can capture a specific level of patterns, but the potential for improvement is high. Several changes in the performance of the models were observed when only the feature selection strategy was applied. The CNN model significantly improved in this scenario, increasing the validation accuracy to 91.68% and the Kappa value to 0.7516. This shows that CNN can benefit from a more focused feature set and generalize better with reduced noise. For MLP, SVM, and XGBoost, feature selection alone improved some metrics but showed results more sensitive to imbalance, especially for Kappa and MCC. This indicates that feature selection alone cannot solve the class imbalance problem. Comparisons of different models and pre-processing steps during training are given in Figure 6.

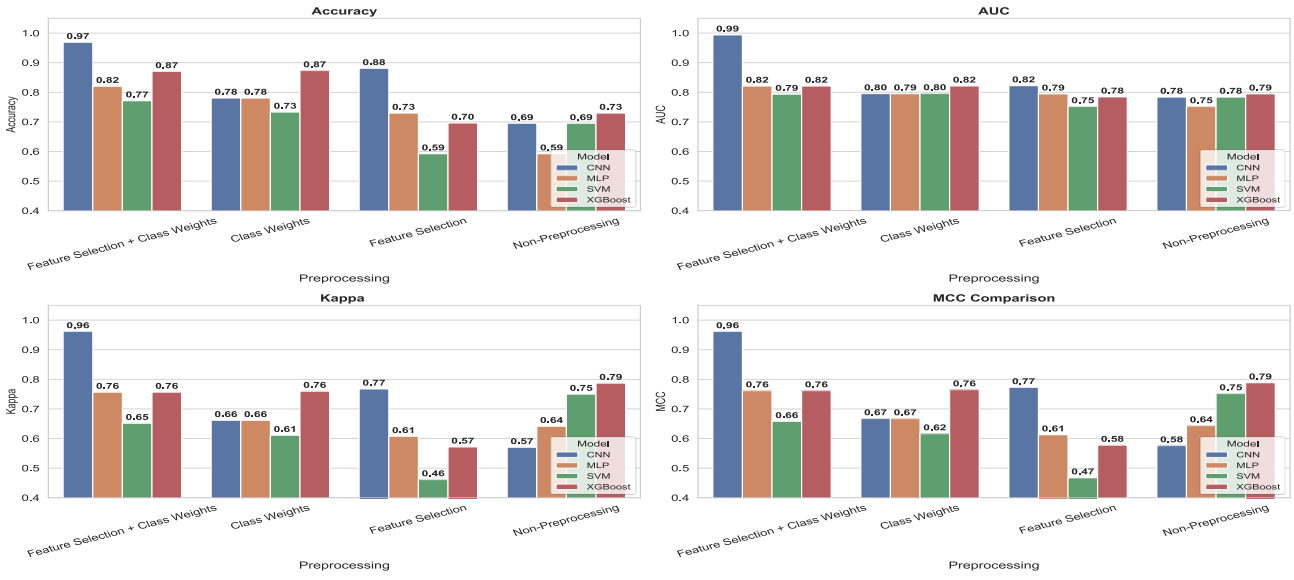

Figure 6. Performance results of different models obtained in training.

Applying only the class weighting strategy significantly increased all models' Kappa and MCC values, especially in the validation set. This indicates that class weighting improves the model's overall predictive stability and reliability by encouraging better learning of minority classes in the dataset. Therefore, it shows that class imbalance in the data is an essential step in classification performance, and the class weighting method applied overcomes this challenge. In the case of class weighting and feature selection, the most comprehensive pre-processing approach and the most fundamental components of the DeepInsulin-Net method, the models showed the highest performance. Under this combined strategy, CNN achieved extremely high and balanced performance metrics such as 94.03% validation accuracy, 0.9921 AUC, 0.9259 Kappa and 0.9259 MCC. On a model-by-model basis, the proposed CNN architecture exhibited a clear superiority over all other tested algorithms, especially when combined with pre-processing strategies. CNN's high training set performance shows that the model can learn data patterns efficiently. In contrast, the high metrics in the validation set indicate a strong generalization capability. XGBoost performed quite competitively and was closest to CNN in all scenarios, especially when class weighting was included. MLP yielded acceptable results, especially with combined pre-processing, while SVM generally underperformed compared to the other models on this dataset and problem definition.

6. Conclusion

This study presents a deep learning-based approach to predicting clinical side effects caused by drug-drug interactions. The proposed method uses molecular fingerprinting and physicochemical properties of drugs to more precisely model the adverse impact of interactions. Seventeen common clinical side effects associated with insulin are selected from the TWOSIDE dataset using a text mining method. Drug pairs were chosen from the dataset, and their molecular properties were calculated using MACCS, Morgan fingerprints and RDKit descriptors. These calculated features using SMILES strings of drug molecules were combined into a single matrix format for each drug pair. In this step, the Variance Thresholding method was applied to the attribute matrix to select the more significant attributes for classification. Drug pairs were used as inputs to a 1D CNN model designed as parallel inputs to predict drug interactions. In the parallel model intended to classify the interactions with high performance, the calculated attributes of the drug pairs are applied to the model as input from different layers, and these attribute matrices are combined in the deep layers of the model and fed to the output layer. The proposed method uses class weighting to balance the unbalanced class distribution. Experimental results show that the proposed model accurately predicts clinical side effects. The model reaches 99.66% accuracy in the training phase, while the validation accuracy is 96%. The training loss decreased to 0.01, while the validation loss stabilized at 0.22. The ROC-AUC metric remained above 0.99 throughout all epochs, and the model demonstrated a strong generalization capability, especially in detecting rare adverse events. These results show that the model can predict common and specific clinical side effects successfully. The developed deep learning architecture allows the features of each drug to be processed independently and combined to predict the probability of occurrence of a specific clinical side effect caused by a drug-drug interaction. This approach increases the model's scalability and reduces the computational burden. The proposed approach can be essential in drug safety assessments and clinical decision support systems. Integrating the model into clinical practice can provide data-driven support for clinicians to make more informed decisions in drug prescribing processes. Implementing these proposed improvements and research directions can significantly improve the current model's performance, robustness and interpretability for predicting adverse drug-drug interactions. Integrating more comprehensive and diverse datasets, applying advanced feature engineering techniques, and exploring innovative model architectures will enable higher accuracy and generalizability, especially in detecting complex and unique interactions. Transparency of the model's decision mechanisms through the integration of explainable artificial intelligence methods will increase clinical acceptability while considering critical factors such as polypharmacy and dose dependency, making predictions more compatible with clinical practice. Future work could develop a more comprehensive solution by integrating the proposed method with different data types, analyzing multiple interacting drug combinations, and its applicability in real-time clinical environments.

References

[1] A. Agustí, E. Melén, D. L. DeMeo, R. Breyer-Kohansal, and R. Faner, “Pathogenesis of chronic obstructive pulmonary disease: understanding the contributions of gene–environment interactions across the lifespan,” *lancet Respir. Med.*, vol. 10, no. 5, pp. 512–524, 2022.

[2] S. R. Chowdhury, D. C. Das, T. C. Sunna, J. Beyene, and A. Hossain, “Global and regional prevalence of multimorbidity in the adult population in community settings: a systematic review and meta-analysis,” *EClinicalMedicine*, vol. 57, 2023.

[3] R. W. Hoel, R. M. G. Connolly, and P. Y. Takahashi, “Polypharmacy management in older patients,” in *Mayo Clinic Proceedings*, Elsevier, 2021, pp. 242–256.

[4] F. Pazan and M. Wehling, “Polypharmacy in older adults: a narrative review of definitions, epidemiology and consequences,” *Eur. Geriatr. Med.*, vol. 12, pp. 443–452, 2021.

[5] J. Wolff *et al.*, “Polypharmacy and the risk of drug–drug interactions and potentially inappropriate medications in hospital psychiatry,” *Pharmacoepidemiol. Drug Saf.*, vol. 30, no. 9, pp. 1258–1268, 2021.

[6] M. R. Mohamed *et al.*, “Association of polypharmacy and potential drug-drug interactions with adverse treatment

outcomes in older adults with advanced cancer," *Cancer*, vol. 129, no. 7, pp. 1096–1104, 2023.

[7] M. Zitnik, M. Agrawal, and J. Leskovec, "Modeling polypharmacy side effects with graph convolutional networks," *Bioinformatics*, vol. 34, no. 13, pp. i457–i466, 2018.

[8] H. Askr, E. Elgeldawi, H. Aboul Ella, Y. A. M. M. Elshaier, M. M. Gomaa, and A. E. Hassanien, "Deep learning in drug discovery: an integrative review and future challenges," *Artif. Intell. Rev.*, vol. 56, no. 7, pp. 5975–6037, 2023.

[9] M. F. Rasool *et al.*, "Assessment of risk factors associated with potential drug-drug interactions among patients suffering from chronic disorders," *PLoS One*, vol. 18, no. 1, p. e0276277, 2023.

[10] P. Rorsman and E. Renström, "Insulin granule dynamics in pancreatic beta cells," *Diabetologia*, vol. 46, pp. 1029–1045, 2003.

[11] N. J. Hogrebe, K. G. Maxwell, P. Augsornworawat, and J. R. Millman, "Generation of insulin-producing pancreatic β cells from multiple human stem cell lines," *Nat. Protoc.*, vol. 16, no. 9, pp. 4109–4143, 2021.

[12] M. A. Hill *et al.*, "Insulin resistance, cardiovascular stiffening and cardiovascular disease," *Metabolism*, vol. 119, p. 154766, 2021.

[13] M. S. Rahman *et al.*, "Role of insulin in health and disease: an update," *Int. J. Mol. Sci.*, vol. 22, no. 12, p. 6403, 2021.

[14] A. Kleinridders, H. A. Ferris, W. Cai, and C. R. Kahn, "Insulin action in brain regulates systemic metabolism and brain function," *Diabetes*, vol. 63, no. 7, pp. 2232–2243, 2014.

[15] J. J. Holst, L. S. Gasbjerg, and M. M. Rosenkilde, "The role of incretins on insulin function and glucose homeostasis," *Endocrinology*, vol. 162, no. 7, p. bqab065, 2021.

[16] J. Blahova, M. Martiniakova, M. Babikova, V. Kovacova, V. Mondockova, and R. Omelka, "Pharmaceutical drugs and natural therapeutic products for the treatment of type 2 diabetes mellitus," *Pharmaceuticals*, vol. 14, no. 8, p. 806, 2021.

[17] R. A. Haeusler, T. E. McGraw, and D. Accili, "Biochemical and cellular properties of insulin receptor signalling," *Nat. Rev. Mol. cell Biol.*, vol. 19, no. 1, pp. 31–44, 2018.

[18] G. Wilcox, "Insulin and insulin resistance," *Clin. Biochem. Rev.*, vol. 26, no. 2, p. 19, 2005.

[19] D. Soumya and B. Srilatha, "Late stage complications of diabetes and insulin resistance," *J Diabetes Metab*, vol. 2, no. 9, p. 1000167, 2011.

[20] O. I. Aruoma, V. S. Neergheen, T. Bahorun, and L.-S. Jen, "Free radicals, antioxidants and diabetes: embryopathy, retinopathy, neuropathy, nephropathy and cardiovascular complications," *Neuroembryology Aging*, vol. 4, no. 3, pp. 117–137, 2007.

[21] C. Argano, L. Mirarchi, S. Amodeo, V. Orlando, A. Torres, and S. Corrao, "The role of vitamin D and its molecular bases in insulin resistance, diabetes, metabolic syndrome, and cardiovascular disease: state of the art," *Int. J. Mol. Sci.*, vol. 24, no. 20, p. 15485, 2023.

[22] J.-M. Guettier and P. Gorden, "Hypoglycemia," *Endocrinol. Metab. Clin.*, vol. 35, no. 4, pp. 753–766, 2006.

[23] L. Sacchetta *et al.*, "Synergistic effect of chronic kidney disease, neuropathy, and retinopathy on all-cause mortality in type 1 and type 2 diabetes: a 21-year longitudinal study," *Cardiovasc. Diabetol.*, vol. 21, no. 1, p. 233, 2022.

[24] B. T. Alemu, O. Olayinka, H. A. Baydoun, M. Hoch, and M. Akpinar-Elci, "Neonatal hypoglycemia in diabetic mothers: A systematic review," *Curr. Pediatr. Res.*, vol. 21, no. 1, 2017.

[25] T. Zhang, J. Leng, and Y. Liu, "Deep learning for drug–drug interaction extraction from the literature: a review," *Brief. Bioinform.*, vol. 21, no. 5, pp. 1609–1627, 2020.

[26] 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, 2020.

[27] K. Han *et al.*, "A review of approaches for predicting drug–drug interactions based on machine learning," *Front. Pharmacol.*, vol. 12, p. 814858, 2022.

[28] 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, 2023.

[29] A. Akgul, Y. Karaca, M. A. Pala, M. E. Çimen, A. F. Boz, and M. Z. Yildiz, "Chaos Theory, Advanced Metaheuristic Algorithms And Their Newfangled Deep Learning Architecture Optimization Applications: A Review," *Fractals*,

vol. 32, no. 03, p. 2430001, 2024.

[30] M. A. Pala and M. Z. Yıldız, "Improving cellular analysis throughput of lens-free holographic microscopy with circular Hough transform and convolutional neural networks," *Opt. Laser Technol.*, vol. 176, p. 110920, 2024.

[31] M. A. Pala, "Graph-Aware AURALSTM: An Attentive Unified Representation Architecture with BiLSTM for Enhanced Molecular Property Prediction," *Mol. Divers.*, Apr. 2025, doi: 10.1007/s11030-025-11197-4.

[32] X. Li, Z. Xiong, W. Zhang, and S. Liu, "Deep learning for drug-drug interaction prediction: A comprehensive review," *Quant. Biol.*, vol. 12, no. 1, pp. 30–52, 2024.

[33] E. Kim and H. Nam, "DeSIDE-DDI: interpretable prediction of drug-drug interactions using drug-induced gene expressions," *J. Cheminform.*, vol. 14, no. 1, p. 9, 2022.

[34] Z. Li, X. Tu, Y. Chen, and W. Lin, "HetDDI: a pre-trained heterogeneous graph neural network model for drug–drug interaction prediction," *Brief. Bioinform.*, vol. 24, no. 6, p. bbad385, Nov. 2023, doi: 10.1093/bib/bbad385.

[35] M. Asfand-E-Yar, Q. Hashir, A. A. Shah, H. A. M. Malik, A. Alourani, and W. Khalil, "Multimodal cnn-ddi: using multimodal cnn for drug to drug interaction associated events," *Sci. Rep.*, vol. 14, no. 1, p. 4076, 2024.

[36] N. P. Tatonetti, P. P. Ye, R. Daneshjou, and R. B. Altman, "Data-driven prediction of drug effects and interactions.," *Sci. Transl. Med.*, vol. 4, no. 125, p. 125ra31, Mar. 2012, doi: 10.1126/scitranslmed.3003377.

[37] A. Cereto-Massagué, M. J. Ojeda, C. Valls, M. Mulero, S. Garcia-Vallvé, and G. Pujadas, "Molecular fingerprint similarity search in virtual screening," *Methods*, vol. 71, pp. 58–63, 2015.

[38] S. Zhong and X. Guan, "Count-based morgan fingerprint: A more efficient and interpretable molecular representation in developing machine learning-based predictive regression models for water contaminants' activities and properties," *Environ. Sci. Technol.*, vol. 57, no. 46, pp. 18193–18202, 2023.

[39] G. Landrum *et al.*, "rdkit/rdkit: 2024_09_5 (Q3 2024) Release." Zenodo, 2025. doi: 10.5281/zenodo.14779836.

[40] X. Ying, "An overview of overfitting and its solutions," in *Journal of physics: Conference series*, IOP Publishing, 2019, p. 22022.

[41] M. A. F. A. Fida, T. Ahmad, and M. Ntahobari, "Variance threshold as early screening to Boruta feature selection for intrusion detection system," in *2021 13th International Conference on Information & Communication Technology and System (ICTS)*, IEEE, 2021, pp. 46–50.

[42] M. A. PALA, M. E. ÇİMEN, M. Z. YILDIZ, G. ÇETİNEL, E. AVCIOĞLU, and Y. ALACA, "CNN-Based Approach for Overlapping Erythrocyte Counting and Cell Type Classification in Peripheral Blood Images," *Chaos Theory Appl.*, 2022, doi: 10.51537/chaos.1114878.

Conflict of Interest Notice

The author declares that there is no conflict of interest regarding the publication of this paper.

Artificial Intelligence Statement

Artificial intelligence tools were used solely for language and grammar correction during the preparation of this manuscript. No AI tools were involved in the research design, data analysis, interpretation of results, or writing of the scientific content. All intellectual contributions and scholarly work were carried out exclusively by the author.

Plagiarism Statement

This article has been scanned by iThenticate ™.