Drug-Drug Interaction Prediction Using Krill Herd Algorithm Based on Deep Learning Method

Abdulsamad Al-Marghilani

College of Computer Science & Information Technology, Northern Border University,
P.O.Box 840, Rafha 91911, Saudi Arabia

Abstract
Parallel administration of numerous drugs increases Drug-Drug Interaction (DDI) because one drug might affect the activity of other drugs. DDI causes negative or positive impacts on therapeutic output. So there is a need to discover DDI to enhance the safety of consuming drugs. Though there are several DDI system exist to predict an interaction but nowadays it becomes impossible to maintain with a large number of biomedical texts which is getting increased rapidly. Mostly the existing DDI system address classification issues, and especially rely on handcrafted features, and some features which are based on particular domain tools. The objective of this paper to predict DDI in a way to avoid adverse effects caused by the consumed drugs, to predict similarities among the drug, Drug pair similarity calculation is performed. The best optimal weight is obtained with the support of KHA. LSTM function with weight obtained from KHA and makes bets prediction of DDI. Our methodology depends on (LSTM-KHA) for the detection of DDI. Similarities among the drugs are measured with the help of drug pair similarity calculation. KHA is used to find the best optimal weight which is used by LSTM to predict DDI. The experimental result was conducted on three kinds of dataset DS1 (CYP), DS2 (NCYP), and DS3 taken from the DrugBank database. To evaluate the performance of proposed work in terms of performance metrics like accuracy, recall, precision, F-measures, AUPR, AUC, and AUROC. Experimental results express that the proposed method outperforms other existing methods for predicting DDI. LSTM-KHA produces reasonable performance metrics when compared to the existing DDI prediction model.

Key words: Drug-Drug Interaction, Long Short Term Memory, Krill Herd Algorithm, Deep Learning, Machine Learning

1. Introduction

Drug-Drug Interaction (DDI) causes pharmacokinetic or pharmacodynamics effects on the human body, when two or more drugs consumed at a time [1,2]. They are the general cause of adverse drug reaction (ADR) and increase the cost of healthcare. The major cause for ADR is due to unplanned DDI and irregularly arise of co-prescription of drugs. Whereas DDI prediction is ideal during a clinical trial, every interaction must be reported once drugs were approved for clinical usage. From the literature survey, it is found that most of the readmission cases in hospitals are happened due to ADR.

Actually, acetylsalicylic acid generally referred to as aspirin, a drug utilized for the treatment of fever and pain because of several causes. This medicine had both antipyretic and anti-inflammatory effects, which obstruct platelet aggregation and utilized for the prevention of myocardial infarction and blood clots. However, the severity or risk of hypertension could be increased when there is a combination of acetylsalicylic acid and 1-benzimidazole [3]. Detection of DDI minimize unanticipated DDI, reduce the drug production cost, and could be utilized to optimize the designing process of the drug. Therefore, there is a need to have knowledge related to both ADR and DDI for clinical applications and the development of drugs, mainly for co-administered medications.

Nowadays several researchers concentrate on ADR which occurs among pairs of drugs. Further, to minimize cost and to analyze possible DDI, an automatic technique is needed to discover ADR. Existing methods involves in clinical examination of drugs and post-marketing surveillance. Some aspects were mined from drug properties then statistical method and machine learning algorithm were implemented to predict DDI but it's not much efficient in predict accurately.
In the upcoming year, deep learning technique [4] has acquired promising result in multiple NLP (Natural Language Processing) tasks like DDI prediction [5] relationship extraction [6], image classification, speech recognition, etc. These methods have a big advantage like easy adaption to various domains. Similarly, big data analytics is an advanced technique used to extract the most useful data from various kinds of large data set to identify user preference, market trends, and hidden patterns for successful decision making. Big Data analytics [7] use sophisticated algorithms depending on deep learning methods to process a large quantity of real-time data with high accuracy and efficiency.

Deep learning has been introduced to automatically learn the feature illustrations from plentiful unannotated data. DDI extractor's features were learned automatically with help of DNN (Deep Neural Network) without manual support. Depending on the NN structure, DDI prediction

Manuscript received June 5, 2021
Manuscript revised June 20, 2021
https://doi.org/10.22937/IJCSNS.2021.21.6.41

model is basically categorized into 2 types: RNN dependent DDI prediction model and CNN dependent the DDI prediction model. According to [8] CNN dependent DDI extraction model used. Researchers use utilized shallow CNN and integrated position embedding with word embedding in the CNN model. Apart from CNN, RNN also has an advantage over-classification of biomedical relations. By [9] RNN based DDI extraction model was utilized with numerous attention layers to classify DDI.

To overcome all the above-said issues, in this paper deep learning dependent method such as LSTM with Krill Herd optimization algorithm is used to extract deep features for detecting DDI. Among all nature inspired algorithm, KHA is a new optimization technique which was inspired by the behavior of krill herd. KHA is employed to resolve various kinds of real-world optimization issues either by merging with others type of evolutionary algorithms to enhance basic KHA or by the addition of few mathematical ideas [10]. LSTM captures globally and locally prominent features from a drug which is most useful for the prediction of DDI.

The objectives of this research are:
(a) To predict DDI in a way to avoid adverse effects caused by the consumed drug.
(b) To predict similarities among the drug, novel Drug pair similarity calculation is performed.
(c) The best optimal weight is obtained with the support of KHA.
(d) LSTM function with weight obtained from KHA and makes the best prediction of DDI.

1.1 Paper Organization

The following section 2 describes the literature review related with proposed study. Section 3 elaborates the proposed methodology LSTM-KHA method. Section 4 illustrates the results and discussion of the proposed study. Finally the paper is concluded in section 5.

2. Related Works

This research, introduced a Drug-Drug Interaction (DDI) prediction method depending on key biological elements like transporters, targets (CTET), and enzymes. This model was used for 2189 accepted drugs. For every drug, all linked CTET was gathered and an equivalent binary vector was built to predict DDI. Several similarity measures were conducted to predict DDI. Among the evaluated similarity approach, inner product dependent similarity measures offer enhanced detection values. Among 2,394,766 drug pairs, the introduced model had the potential to predict 250,000 unknown potential DDI. From the findings, the researcher proposed a method in the silicon approach for predicting DDI. The researcher forecast that this proposed method can be utilized for the prediction of DDI depending on functional similarities of drugs [11].
Three LSTM (Long Short Term Memory) model such as Join AB-LSTM, AB-LSTM, and Bi-LSTM have been presented. These three models utilize position and word embedding as a latent feature, so there is no need for explicit feature engineering. From the sentence, implicit features are extracted with the use of Bi-LSTM model. Whereas Joint AB-LSTM and AB-LSTM use alternative pooling in Bi-LSTM's output layer in a way to allow weight to features. The experiment was conducted on the SemEval-2013 dataset and results express that Joint AB-LSTM offers reasonable results with 69.39% of F-score [12].

This study, proposed a new method for choosing the optimal weight for HES (hybrid Energy System). The reason was to reduce HES's total cost for various constraints like capacity restriction for battery and charge restriction for a fuel cell. Issues related to optimization had been resolved with use of enhanced Krill Herd Algorithm (KHA) and converged krill Herd (CKH) algorithm. The simulation was conducted to analyze speed demand and average power demand of locomotive slope and locomotive. The result illustrates that CKH gives a better result than a traditional optimization algorithm [13].

ISCMF (Integrated Similarity Constrained Matrix Factorization) for detecting DDI are presented. 8 similarities were computed depending on drug substructure, side effects, enzymes, off-label side effects, targets, indication data, transporters and pathways, and Gaussian interaction profile for drug pairs. Consequently, a non-linear similarity fusion approach was utilized to combine numerous similarities and create them as informative. Finally, Implemented ISCMF which projects drugs in interaction space to lower rank space to acquire new insight for DDI. Experimental results express that the proposed method attains better results for 5 fold cross-validation. This method enhances AUPR to 10% and F-measure to 18% [14]
AGCN (Attention- dependent Graph Convolutional Network) to resolves the arising issues have been discussed in this study. AGCN was modelled in a way to control structural and contextual knowledge together, where GCN was implemented in integration with encoder dependent recurrent network. In addition, the researcher uses new attention dependent pruning plan to optimally utilize syntactic data by ignoring irrelevant data. So AGCN handles the structure and the context of input sentences more effectively. Experiments were conducted on dominant DDI mining corpus and result express that proposed model outperforms existing models [15].

Bidirectional LSTM method called (Biomedical Resource) BR-LSTM have been developed. This proposed method was a combination of lexical data, biomedical resource, and data related to entity position to mine DDI from the literature of biomedical. Experiments were conducted on SemEval 2013 Task 9.2 dataset to examine

the performance of the proposed method and the result shows that BR-LSTM was superior to other existing methods, which achieves 0.711% of F-score [16]. Furthermore, this study presented neighborhood distance idea and genetic reproduction mechanism in KHA to develop KHAMC (KHA with mutation and crossover) and KHA with the idea of neighborhood distance is referred as KHAMCD. Here the researcher compares the performance of KHA, KHAMC, and KHAMCD. Experiments were conducted on 8 benchmark dataset and 2 real-world ELD (Economic Load Dispatch) issues of a power system. Results express that the KHAMCD method outperforms the other two methods in resolving non-smooth and smooth constrained optimization issues [17].

Novel machine learning method for detecting DDI depending on numerous data sources have been presented. For this task, 12000 drugs are used from PharmGKB, KEGG drug, and DrugBank, all these are merged by utilizing KG (Knowledge Graph). The prediction model was trained by embedding nodes in the graph with the use of several embedding methods. Among several embedding methods, the best one is the ComplEx embedding technique which was developed using PBG (PyTorch-Bigraph) with (Convolutional) C-LSTM and classic machine learning depending on detection models. The proposed method offers better results for F1-score of 0.92, AUPR of 0.94, and MCC of 0.80 during the five-fold-cross-validation test [18]. Cuckoo Search (CSKH) method to resolve issues due to structural optimization has been introduced. CSKH enhances KHA by merging the KA operator coined from the CS algorithm with KHA. In CSKH, a greedy selection policy was utilized by considering originals from CS and KHA. Additionally, in a way to enhance the assessment of CSKH, part of the worst krill was thrown away and replaced with a randomly developed new one with the use of KA operators towards the end of every generation. Experiments were conducted on 5 real engineering issues to validate its performance and the result proves that CSKH had the potential in resolving issues related to constrained engineering models [19].

Quantum behaved PSO (Particle Swarm Optimization) and KH, referred as KH-QPSO for engineering and benchmark optimization have been introduced. QPSO was proposed for improving the capacity of local search and enhance individual diversity in the population. KH-QPSO had the capacity of eliminating premature convergence and minimize finding function and KH-QPSO had made the whole individual to proceeds towards true global optimum without using additional operators to QPSO and KH algorithms. Experiments were conducted to evaluate proposed methods performance and result express that KH-QPSO was more effective than the existing optimization technique for resolving engineering optimization issues and standard test issues [20]. On the other hand, KMR, Knowledge Oriented feature-driven technique which studied drug regarding knowledge with -accurate illustration has been proposed. Experiments were conducted on real-world medical dataset to examine the performance of the proposed method. Result proves that KMR helps in finding DDI with 99.7% of accuracy thereby enhance the quality of prediction of new drugs [21].

Moreover, Examining risk regarding DDI is most important to investigate novel drugs during the development process of the drug. In vitro, several experiments were conducted to assess metabolism and transporter-mediated of DDI to investigate new drugs. These experimental results were inferred with support of IVIVE (*In vitro - In Vivo* Extrapolation) methods, to examine how to estimate DDI clinically for management of DDI plans which includes alternate therapies, recommended dosage, and in a diverse patient population. This paper offers an overview of basic IVIVE method and the existing vitro experimental system for metabolism and transporter-mediated DDI [22]. Further in another research gathers various kinds of drug data which would be useful for the prediction of DDI. For this purpose, 3 models were introduced such as neighbor method, matrix perturbation method, and random walk method to construct the prediction model depending on various data. Additionally, various models with appropriate ensemble rules like a classifier, weight average ensemble rule and develop ensemble design was utilized to attain the best performance [23]. Sparse feature learning ensemble technique with linear neighborhood regularization referred as SFLLN to detect DDI. Primarily, 4 drug features like targets, chemical substructures, pathways and enzymes were integrated by linking drugs in various feature space into common interaction space over space feature learning. Additionally, presented linear neighborhood regularization to explain the relationship among drug-drug in interaction space by utilizing known DDI. Experiments were conducted on a benchmark dataset and result express that SFLLN achieves higher accuracy with a reasonable quantity of running time for predicting DDI [24]. Position-aware deep multi-task method for mining DDI from biomedical text has been proposed. Attention-BiLSTM was utilized to encode every sentence. The position of word with target drug in text was merged with BiLSTM's hidden state to generate position-aware attention weights. Experiments were conducted on DDI Extraction challenge 2013 corpus and result express that the proposed method outperforms all existing methods [25].

A novel deep learning model for more accurate prediction of DDIs and their effects, which may assist in future research to discover novel DDIs and their pharmacological effects have been presented [26]. KHA is employed to resolve various kinds of real-world optimization issues either by merging with others type of evolutionary algorithms to enhance basic KHA or by the addition of few mathematical ideas [10]. LSTM captures

globally and locally prominent features from a drug which is most useful for the prediction of DDI. To overcome all the above-said issues, in this paper deep learning dependent method such as LSTM with Krill Herd optimization algorithm is used to extract deep features for detecting DDI.

3. Materials and Methods

In the article, LSTM is combined with the KHA optimization algorithm to accurately predict DDI to avoid ADR. To evaluate the performance of the proposed method DrugBank dataset is used.

3.1 Input Dataset

DS1 (CYP) and DS2 (NCYP) [27] dataset which includes drugs 807, 7 kinds of similarity matrix in which 4 depends on ATC, ligand-dependent chemical similarity, chemical similarity and side effects. 3 more similarity measures were built on drug target similarities like sequence similarity, distance on PPI (Protein-Protein Interaction) network, and GO annotations. All these data are gathered from DrugBank database.

DrugBank is the most credible database for known DDI [3] which includes more than 300000 DDIs. Yet, this quantity of drug interaction information is less than 1% of the total quantity of drug pairs which exist in DrugBank. Knowledge regarding pharmaceutical formulation, drug target, physiological effect, and drug chemical structure are studied from Drug Bank (DS3) which includes drug entries of 11680 which includes 2625 accepted small molecule drug, accepted 1115 biotech drugs, nutraceuticals 128 over 5504 experimental drugs [21]. Drug class aspects were mined from ChemOnt [21] which is computable, and comprehensive chemical taxonomy with annotated chemical ontology which follows cheminformaticians to do automatic, and rapid chemical classification. [28] presented an approach the drug-drug interaction prediction problem as a link prediction problem and present two novel methods for drug-drug interaction prediction based on artificial neural networks and factor propagation over graph node.

3.2 Pre-processing

Standard pre-processing steps are carried out on training, and testing data. These step consist of sentence segmentation, tokenization, PSO (Part of Speech) tagging, and syntactic parsing which develop the dependency graph, and constituent parse tree for sentence. To guarantee feature generalization, the anonymous drugs is referred as “DRUG” for target drug and for other drug is referred as DRUG_OTHER. Numbers were substituted with generic tag “NUM” and other tokens were normalized to their equivalent lemmas by Bio-Lemmatizer.

In some cases, the same drug name might appear for numerous times in a sentence. Where drugs are unlikely to interact with themselves so there is a need to remove the same name drug. This minimizes candidate drug pairs. In some situation, drug name was separated by a colon from a brief explanation of their interaction with other drugs. For example, “Morphine: Association of hormonal contraceptives might improve clearance of morphine”. In such instance, the description is single

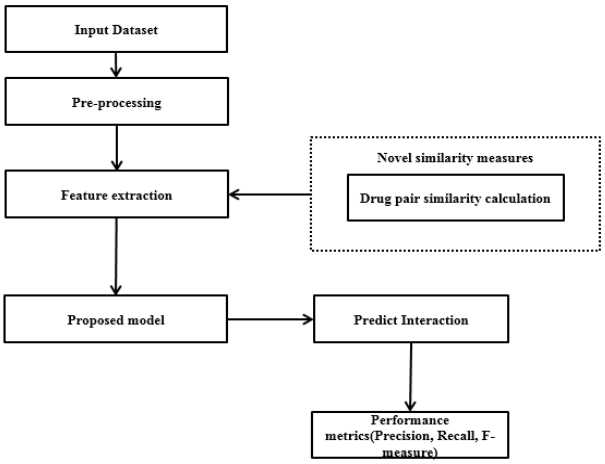


Figure 1: Overall Flow of Proposed Method

sentence, pairing “Morphine” on left side of colon with mentioned drug name in explanation might interfere with description’s flow. For further process, drug mentioned on left side of colon is removed.

3.3 Feature Extraction

Weight of every edge is computed as per equation 1, where e denotes edge’s step length and c denotes distance between y and x in referred and referring sequences.

$$w(y,x) = -\frac{1}{e^{1+c}} \qquad (1)$$

Similarity among referred and referring sequence is computed as summation of all weight,

$$\text{Similarity} = \sum_{\forall x,y} W/2 \qquad (2)$$

3.4 Proposed Model

3.4.1 Krill Herd Algorithm (KHA)

KHA is a swarm intelligence optimization algorithm which depends on krill herd’s behavior utilized to enhance the efficiency of the LSTM network. In KHA, krill searches

for sources of food in multidimensional space and then various decisions were proposed. However, the target is the distance among individual krill and food which is linked with the cost. Accordingly, time-based location of individual krill is computed with the use of 3 operational processes such as: 1). Movement induced by the existence of other individuals. 2). Foraging motion and 3). Random physical diffusion [29,30,31,32,33]. The significant advantage of KHA is: every agent plays a role in the process, derivative data is not needed, it considers of advantages of mutation and crossover operators. On the other side, there is a need for an optimal method to predict initial krill parameter and distribution and basic movement in KHA.

a. Movement induced by existence of other individuals

Individual krill's speed is influenced by other individual krill's movement in multidimensional space, where speed varies dynamically and drastically by target herd, local impact, and repulsive impact. [30] Motion of individual krill are shown as:

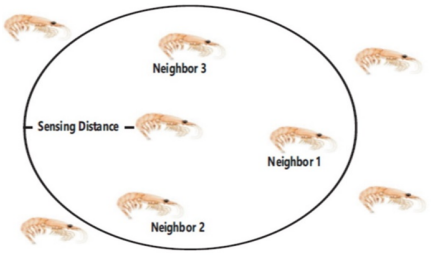


Figure 2: Structure of herd sensing around individual krill. Adapted from [30].

$$\alpha_j^{new} = \varepsilon_j \alpha_j^{max} + \mu_m \alpha_j^{old} \quad (3)$$

$$\varepsilon_j = \varepsilon_j^{local} + \varepsilon_j^{ideal} \quad (4)$$

$$\varepsilon_j^{local} = \sum_{j=0}^{Ks-1} \kappa_{ji} * x_{ji} \quad (5)$$

$$\kappa_{ji} = {}^{\kappa_j - \kappa_i}/_{\kappa_w - \kappa_b} \quad (6)$$

$$x_{ji} = {}^{x_j - x_i}/_{|\kappa_w - \kappa_b| r(0.1)} \quad (7)$$

$$\varepsilon_j^{ideal} = s\left(r(0.1) + {}^{i}/_{i_{max}}\right) \kappa_j^{best} x_j^{best} \quad (8)$$

In the above equation, α_j^{max} denotes maximum induced movement, α_j^{old} represents induced movement. μ_m denotes induced motion's algebraic size, where target impacts are expressed as ε_j^{local} and ε_j^{ideal}. Besides κ_w denotes worst population location and κ_b denotes best population location then κ_j and κ_i denotes fitness value of j^{th} and i^{th}individual krill. Finally imax express maximum and current quantities [30] These sensing distance parameter were utilized to discover every individual krill's neighbors as expressed in fig 2. If distance among 2 individual krill is shorter than sensing distance, then that individual krill is taken as neighbor of the individual krill

$$Sensing\ distance = 1/5k_p \sum_{i=0}^{k_p - 1} |x_i - x_j| \quad (9)$$

b. Foraging Motion

Foraging motion of every individual krill is formulated under conditions of current location of food and earlier knowledge of food location [30].

$$E_n = U_e b_j + \mu_e E_n^{old} \quad (10)$$

$$b_j = b_j^{food} + b_j^{best} \quad (11)$$

Where E_n referes to primary motion, U_e refers to the aging speed and μ_e denotes the algebraic size of foraging in the range [0,1]. Additionally, E_n^{old} represents earlier foraging movement, b_j^{food} refers to food absorption and b_j^{best} express the best fitness for every individual krill [30].

c. Random Physical Diffusion

In KHA, increase in population diversity is done by diffusion function, which is combined with individual krill. Mathematical equation is expressed as:

$$SC_j = SC^{max} u \quad (12)$$

Where SC^{max} illustrates the highest diffusion speed, and u refers to random vector in the range [-1,1].

d. Calculation of Crossover and Mutation operator

Crossover Operator

Every individual krill location is an upgraded based on crossover probability. Updating procedure of i^{th} component of j^{th} krill might be explained as:

$$y_{j,i} = \begin{cases} y_{s,i} \ if \ rand \leq C \\ y_{j,i} \ if \ rand \leq C \end{cases} \text{ where s = 1,2,3,…M} \quad (13)$$

$$C = 0.2 \times e_j^{best} \quad (14)$$

There is a decrease in Crossover probability with an increase in fitness for global best solution C = 0.

Mutation Operator

$$y_{j,i} = y_{best,i} + E\,(y_{1,i} - y_{2,i}),\ E \in (0,1) \quad (15)$$

With support of mutation probability Q_n the revised value is chosen as:

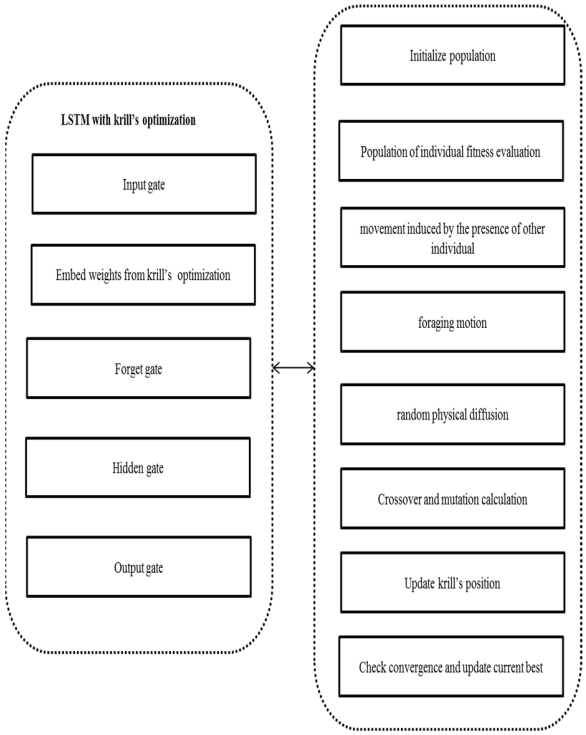


Figure 3: Overall flow of LSTM-KHA

$$y_{j,i}^{new} = \begin{cases} y_{s,i} \ if\ rand \leq Q_n \\ y_{j,i} \ if\ rand \leq Q_n \end{cases} \text{where s = 1,2,3….M} \quad (16)$$

And $Q_n = 0.05 / e_j^{best}$ (17)

e. Mechanism of updating krill's location

During the optimization process, krill alter its location frequently in multidimensional space directed by movements induced by other individual krill, foraging movement and random physical diffusion. All movements function simultaneously, and create algorithms as most robust. Mechanism of upgrading j[th] krill could be illustrated as:

$$y_j^{p+1} = y_j^p + (u_j^p + u_{ej}^p + u_{cj}^p) \times \eta \times \sum_{i=1}^{N} (va_j - la_i) \quad (18)$$

Where va_j denotes upper limit and la_i represents lower limit for i[th] variable, N denotes variable's total count and $\eta \in (0,2)$ is constant.

Every individual krill's location is updated based on their fitness, stop as global optimal solution.

3.4.2 LSTM

Recurrent Neural network (RNN) is a various form of earlier feedforward NN. It is sequence-dependent design, which had the capacity to establish a temporal relationship among recent circumstances and previous data. Whereas RNN is not suitable for time series issues, load forecasting issues for individual households which consumes more energy. Similarly, studying long-range dependencies with RNN is complex because of the gradient vanishing issue. To overcome all the above-described problem, here, LSTM were presented with forgetting gate, hidden gate and output gate. According to [34] naming convention is performed as: let $\{ y_1, y_2, \ldots . y_K \}$ represents LSTM's input sequence, $y_k \in S^t$ denotes t[th] dimension vector of real value at k[th] time step. In a way to create a temporary link, LSTM describes and sustain internal memory cell state for the entire life cycle, which is a significant element of LSTM. Memory cell state s_{k-1} interacts with g_{k-1} intermediate output and the following input y_k to find which element of internal state vector must be maintained, upgraded, or deleted depending on the result of the previous time step and input of the current time step. Additionally, LSTM also describe h_k input node, j_k input gate, e_k forget gate, w_k output gate. Node's formulation in LSTM is as follows:

$$f_g = \sigma\,(\omega_{fy}^{ideal} y_k + \omega_{fT}^{ideal} T_{t-1} + b_f) \quad (19)$$

$$i_g = \sigma\,(\omega_{jy}^{ideal} y_k + \omega_{jT}^{ideal} T_{t-1} + b_j) \quad (20)$$

$$ip_n = \phi\,(\omega_{jpy}^{ideal} y_k + \omega_{jpT}^{ideal} T_{t-1} + b_{jp}) \quad (21)$$

$$o_g = \sigma\,(\omega_{jy}^{ideal} y_k + \omega_{oT}^{ideal} T_{t-1} + b_o) \quad (22)$$

$$m_c = ip_n \odot i_g + m_{c-1} \odot f_g \quad (23)$$

$$T_t = \phi\,(m_c) \odot o_g \text{ f} \quad (24)$$

Where ω_{fy}^{ideal}, ω_{jy}^{ideal}, ω_{jpy}^{ideal}, ω_{jy}^{ideal}, ω_{fT}^{ideal}, ω_{jT}^{ideal}, ω_{jpT}^{ideal}, ω_{oT}^{ideal} denotes the weight matrix for the equivalent input of network activation function, $\odot$ states for element-wise multiplication, σ denotes sigmoid activation function, ϕ denotes tanh function. To train LSTM. There is

a need to state the hyper parameter of hidden output dimension m. In this situation, a hidden output at the provided time step $g_k \in S^m$, which is m-dimensionl vector, so as to r_k.

Generally, g_k and r_k is initialized with 0 tha is g_0 =0 and r_0 =0. There is 3 sigmoid functions, with range for output from 0 to 1, "soft" switch in LSTM is utilized to decide which signal is granted permission to pass gates. If 0 is value for gate, then gate blocks the signal. The decision for e, j and w depend on present y_k input and earlier output g_{k-1}. Input gate's signal controls what to reserve in internal state, forget gate's signal control what to forget from the earlier state r_{k-1}. With updated information of internal state, the output gate decides which r_k internal state must pass as LSTM output g_k. This process is repeated continuously. It's found that biases and weight reduces variation among real training samples and LSTM output. Through this process, data of the present time step is stored and preserved to affect the output of LSTM of a future time step.

4. results and discussion

To determine the efficiency of the proposed work, the proposed method is compared with existing methods on DS1 (CYP) and D S2 (NCYP) [27] dataset which includes drugs 807, 7 kinds of similarity matrix in which 4 depends on ATC, ligand-dependent chemical similarity, chemical similarity and side effects. 3 more similarity measures were built on drug target similarities like sequence similarity, distance on PPI (Protein-Protein Interaction) network and GO annotations. All these data are gathered from the Drug Bank database.

a. Accuracy

It measures the amount of total count of correct categorization:

$$\text{Accuracy} = \frac{True\ Negative\ +True\ Positive}{True\ Negative+False\ Negative+True\ Positive+False\ Positive} \quad (25)$$

b. Precision

It implies a count of correct categorization that is controlled by incorrect categorization.

$$\text{Precision} = \frac{True\ Positive}{False\ Positive\ +True\ Positive} \quad (26)$$

c. Recall

IT measures the count of correct categorization which is taken with a count off missed entries.

$$\text{Recall} = \frac{True\ Positive}{False\ Negative+True\ Positive} \quad (27)$$

d. F-Score

Computes harmonic mean of recall and precision, which provides received effectiveness measurement.

$$\text{F1-score} = 2 * \frac{Recall * Precision}{Recall+Precision} \quad (28)$$

Since all values like accuracy, precision, recall, and F-measures depend on the value of the threshold. There is a need to evaluate the proposed method through AUC (Area Under Curve) where refers to Receiver Operating haracteristics (ROC). Where AUPR refers to Area Under Precision-Recall curve. These metrics denote LSTM-KHA's efficiency which doesn't depend on threshold value. Where AUPR is a fairer criterion used to examine fractions of negative and positive samples.

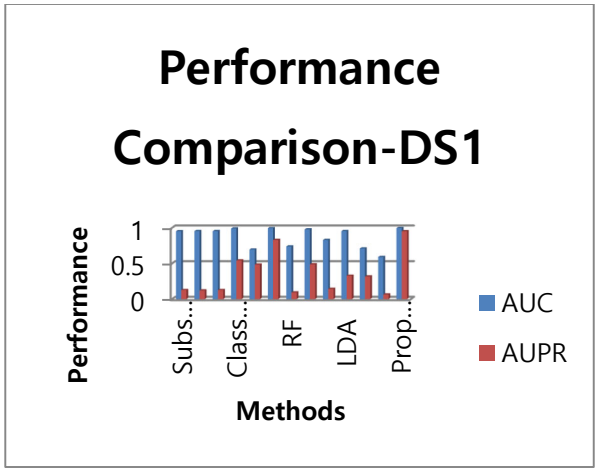


Figure 4: of Performance for existing and LSTM-KHA method on DS1 (CYP) Adapted from [35].

Calculated performance metrics for results of LSTM-KHA in detecting CYP (Cytochrome P450 involved DDI) interactions are expressed in figures 4 and 5, the highest performance in terms of criteria was attained by LSTM-KHA. Proposed LSTM-KHA is most superior to graph-dependent technique. Other existing method fails to provide the expected result. The performance of LSTM-KHA in terms of F-measures, AUPR, precision, AUC, and recall outperforms all existing methods. Even though the ensemble method's performance is not nearer to LSTM-KHA. LSTM-KHA attains 0.996% for AUC and 0.95% for AUPR, this states that LSTM-KHA functions better in detecting CYP DDI. Striking performance of LSTM-KHA on DS1 represents complex relations in the dataset, LSTM-KSH is promising in mining discriminant features which simply finds hidden DDI where other system fails to perform.

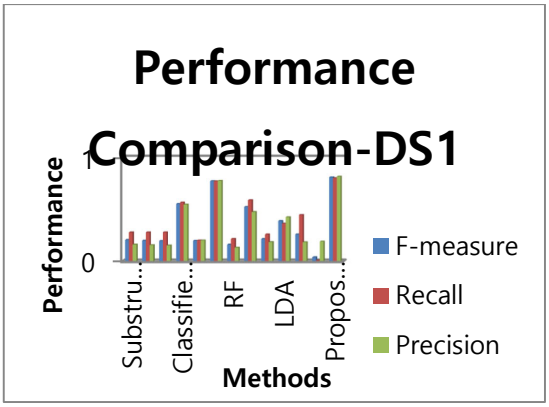


Figure 5: Comparison of Performance for existing and LSTM-KHA method on DS1 (CYP). Adapted from [35].

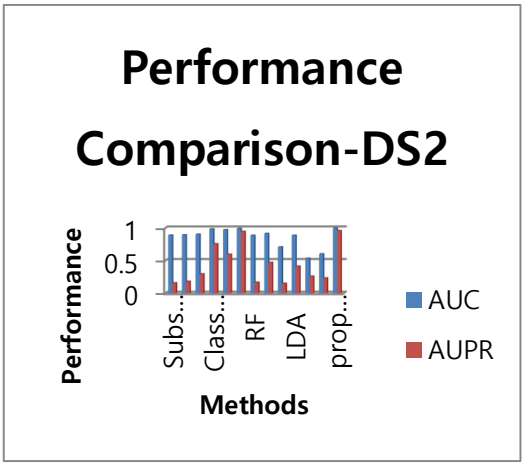


Figure 6: Comparison of Performance for existing and proposed method on DS2 (NCYP). Adapted from [35].

Calculated performance metrics for results of LSTM-KHA in detecting NCYP (DDI without involving Cytochrome P450) interactions are expressed in figures 6 and 7, the highest performance in terms of criteria was attained by LSTM-KHA. The performance of LSTM-KHA in terms of F-measures, AUPR, precision, AUC, and recall outperforms all existing methods. LSTM-KHA acquire the best result for AUC in NCYP DDI detection. Whereas AUC attains 0.997% and 0.958% for AUPR. LSTM-KHA has proved its capability in detecting NCYP DDI.

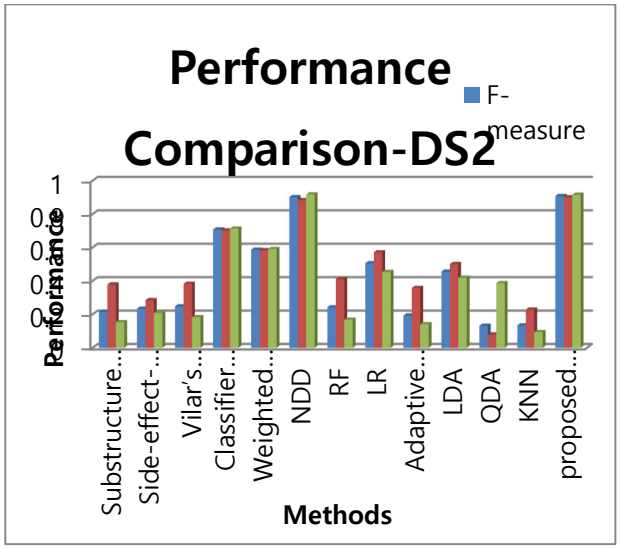


Figure 7: Comparison of Performance for existing and LSTM-KHA method on DS2 (NCYP). Adapted from [35].

To determine the efficiency of the proposed work, the LSTM-KHA is compared with existing methods on (DrugBank dataset) DS3 which includes drug entries of 11680 which includes 2625 accepted small molecule drug, accepted 1115 biotech drugs, nutraceuticals 128 over 5504 experimental drugs[20]. The result expresses varying AUROC and F-Score. There are also numerous factors observed in figures 8 and 9. Compare with other methods, LSTM-KHA enhance the performance of DDI detection. It outperforms the existing system by F-score of 10%, especially because of high precision. ii). High performing method of the existing system in figures 8 and 9, performs depending on several features which were acquired from medical resources. iii). Compared with the existing system, the advantage of LSTM-KHA is that it won't utilize manually distinct features. Features utilized in LSTM-KHA have learned automatically during the training process and consist of most useful data beyond the features which were defined manually. iv) Commonly, all features contribute and create larger performance boosting for detecting DDI. LSTM-KHA outperforms all the tests in terms of performance metrics for DDI prediction.

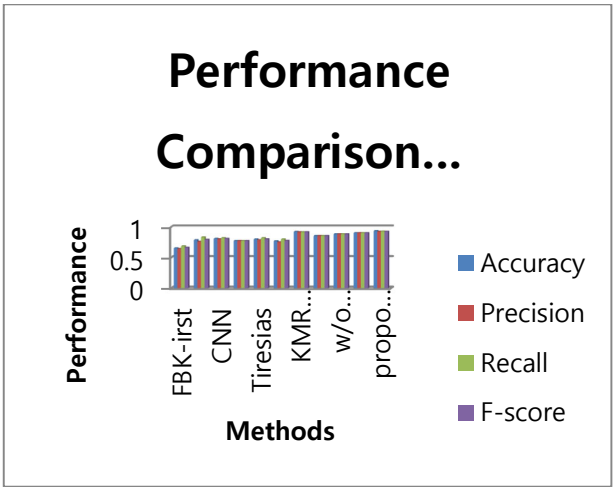


Figure 8: Comparison of Performance for existing and LSTM-KHA method on DS3. Adapted from [21].

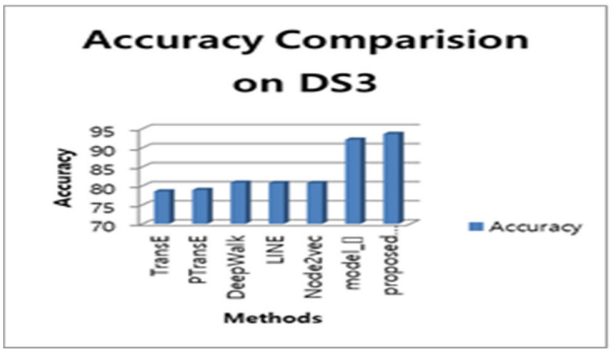


Figure 9 expresses experimental output for the detection LSTM-KHA method on DS3. Adapted from [21].

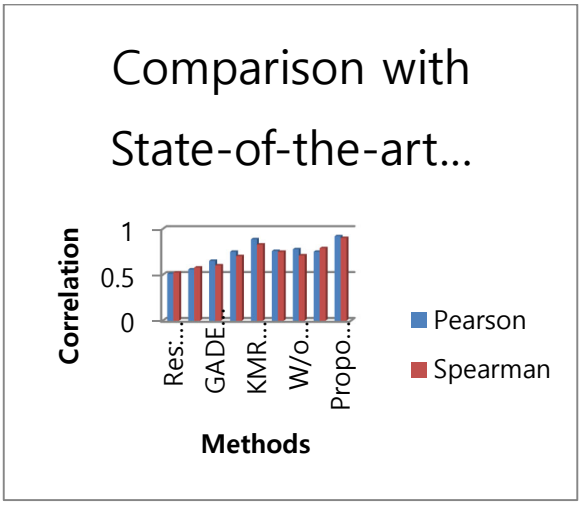


Figure 10: Comparison of state-of-art similarity metrics for existing and LSTM-KHA method on DS3. Adapted from [21].

Figure 10 dipicts experimental result of LSTM-KHA on DrugBank dataset (DS3). Here 4 state-of-the-art were utilized for comparison. The result shows that LSTM-KHA excels all the existing methods with respect to various correlations. LSTM-KHA attain 0.92% for Pearson, and 0.9% for spearman.

5. Conclusion

The LSTM-KHA based DDI prediction model is proposed which was designed with the purpose to overcome existing issues with the DDI prediction model, especially classification issues with drugs. Proposed LSTM-KSHA excels all existing methods in accurately predicting DDI in a way to avoid adverse effects caused by the drugs. Result: Experimental result was conducted on three kinds of dataset DS1 (CYP), DS2 (NCYP) and DS3 taken from DrugBank database. To evaluate the performance of proposed work in terms of performance metrics like accuracy, recall, precision, F-measures, AUPR, AUC and AUROC. Experimental result express that proposed method for DS1 dataset obtains 0.6%for AUC, 12.1% for AUPR, 3.6% for F-score, 21.2% for recall and 27% for precision and DS2 dataset obtains 0.5% for AUC, 1.1% for AUPR, 0.6% for F-score, 1.7% for recall and DS3 dataset obtains 1.41% for Accuracy, 1.17% for AUROC, 2.1% for precision, 2.58% for recall, 2.27% for F-score and 2% for AUPR. These results outperform other existing methods for predicting DDI. LSTM-KHA produces reasonable performance metrics when compared to the existing DDI prediction model.

References

[1] Takeda, T., Hao, M., Cheng, T., Bryant, S. H., & Wang, Y. (2017). Predicting drug–drug interactions through drug structural similarities and interaction networks incorporating pharmacokinetics and pharmacodynamics knowledge. Journal of cheminformatics, 9(1), 16.

[2] Wang, M., Chen, Y., Qian, B., Liu, J., Wang, S., Long, G., & Wang, F. (2017). Predicting Rich Drug-Drug Interactions via Biomedical Knowledge Graphs and Text Jointly Embedding. arXiv preprint arXiv:1712.08875.

[3] Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., . . . Sayeeda, Z. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic acids research, 46(D1), D1074-D1082.

[4] Lamurias, A., Sousa, D., Clarke, L. A., & Couto, F. M. (2019). BO-LSTM: classifying relations via long short-term memory networks along biomedical ontologies. BMC bioinformatics, 20(1), 1-12.

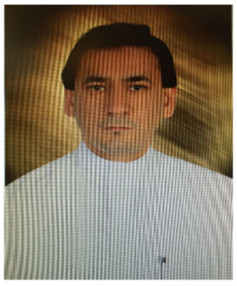
[5] Deng, Y., Xu, X., Qiu, Y., Xia, J., Zhang, W., & Liu, S. (2020). A multimodal deep learning framework for predicting drug-drug interaction events. Bioinformatics.

[6] Lamurias, A., & Couto, F. M. (2019). Text mining for bioinformatics using biomedical literature. Encyclopedia of bioinformatics and computational biology, 1, 602-611.

[7] Jan, B., Farman, H., Khan, M., Imran, M., Islam, I. U., Ahmad, A., Jeon, G. (2019). Deep learning in big data Analytics: A comparative study. Computers & Electrical Engineering, 75, 275-287.

[8] Liu, S., Chen, K., Chen, Q., & Tang, B. (2016). Dependency-based convolutional neural network for drug-drug interaction extraction.

Paper presented at the 2016 IEEE international conference on bioinformatics and biomedicine (BIBM).
[9] Yi, Z., Li, S., Yu, J., Tan, Y., Wu, Q., Yuan, H., & Wang, T. (2017). Drug-drug interaction extraction via recurrent neural network with multiple attention layers. Paper presented at the International Conference on Advanced Data Mining and Applications.
[10] Wang, G.-G., Deb, S., Gandomi, A. H., & Alavi, A. H. (2016). Opposition-based krill herd algorithm with Cauchy mutation and position clamping. Neurocomputing, 177, 147-157.
[11] Ferdousi,R, Safda,i, R, Omidi, Y. (2017) Computational prediction of drug-drug interactions drugs functional similarities, Journal of Biomedical Informatics,V.70, 54- 64.
[12] Sahu, S. K., & Anand, A. (2018). Drug-drug interaction extraction from biomedical texts using long short-term memory network. Journal of biomedical informatics, 86, 15-24.
[13] Guo, Y., Dai, X., Jermsittiparsert, K., & Razmjooy, N. (2020). An optimal configuration for a battery and PEM fuel cell-based hybrid energy system using developed Krill herd optimization algorithm for locomotive application. Energy Reports, 6, 885-894.
[14] Rohani, N., Eslahchi, C., & Katanforoush, A. (2020). ISCMF: Integrated similarity-constrained matrix factorization for drug–drug interaction prediction. Network Modeling Analysis in Health Informatics and Bioinformatics, 9(1), 1-8.
[15] Park, C., Park, J., & Park, S. (2020). AGCN: Attention-based Graph Convolutional Networks for Drug-Drug Interaction Extraction. Expert Systems with Applications, 113538.
[16] Xu, B., Shi, X., Zhao, Z., & Zheng, W. (2018). Leveraging biomedical resources in bi-lstm for drug-drug interaction extraction. IEEE Access, 6, 33432-33439.
[17] Agrawal, Pandit, & Dubey (2016), Improved Krill Herd Algorithm with Neighborhood Distance Concept for Optimization, International Journal of Intelligent Systems and Applications, , PP.34-50.
[18] Rezaul Karim, M., Cochez, M., Bosco Jares, J., Uddin, M., Beyan, O., & Decker, S. (2019). Drug-Drug Interaction Prediction Based on Knowledge Graph Embeddings and Convolutional-LSTM Network. arXiv preprint arXiv:1908.01288.
[19] Abdel-Basset, M., Wang, G.-G., Sangaiah, A. K., & Rushdy, E. (2019). Krill herd algorithm based on cuckoo search for solving engineering optimization problems. Multimedia Tools and Applications, 78(4), 3861-3884.
[20] Li, L., Sun, L., Xue, Y., Li, S., Huang, X. and Mansour, R.F., 2021. Fuzzy Multilevel Image Thresholding Based on Improved Coyote Optimization Algorithm. IEEE Access, 9, pp.33595-33607.
[21] Shen, Y., Yuan, K., Yang, M., Tang, B., Li, Y., Du, N., & Lei, K. (2019). KMR: knowledge-oriented medicine representation learning for drug–drug interaction and similarity computation. Journal of cheminformatics, 11(1), 22.
[22] Yoshida, K., Zhao, P., Zhang, L., Abernethy, D. R., Rekić, D., Reynolds, K. S., . . . Huang, S.-M. (2017). In Vitro–In Vivo Extrapolation of Metabolism-and Transporter-Mediated Drug–Drug Interactions—Overview of Basic Prediction Methods. Journal of pharmaceutical sciences, 106(9), 2209-2213.
[23] Zhang, W., Chen, Y., Liu, F., Luo, F., Tian, G., & Li, X. (2017). Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. BMC bioinformatics, 18(1), 18.
[24] Zhang, W., Jing, K., Huang, F., Chen, Y., Li, B., Li, J., & Gong, J. (2019). SFLLN: a sparse feature learning ensemble method with linear neighborhood regularization for predicting drug–drug interactions. Information Sciences, 497, 189-201.
[25] Zhou, D., Miao, L., & He, Y. (2018). Position-aware deep multi-task learning for drug–drug interaction extraction. Artificial intelligence in medicine, 87, 1-8.
[26] Lee, Park & Ahn, (2019), Novel deep learning model for more accurate prediction of drug-drug Interaction effects, BMC Bioinformatics, https://doi.org/10.1186/s12859-019-3013-0.
[27] Gottlieb, A., Stein, G. Y., Oron, Y., Ruppin, E., & Sharan, R. (2012). INDI: a computational framework for inferring drug interactions and their associated recommendations. Molecular systems biology, 8(1).
[28] Shtar, Rokach & Shapira (2019), Detecting drug-drug interactions using artificial neural networks and classic graph similarity measures, PLoS ONE 14(8): e0219796. https://doi.org/10.1371/journal.pone.0219796.
[29] Abualigah, L. M., Khader, A. T., Hanandeh, E. S., & Gandomi, A. H. (2017). A novel hybridization strategy for krill herd algorithm applied to clustering techniques. Applied Soft Computing, 60, 423-435.
[30] Bolaji, A. L. a., Al-Betar, M. A., Awadallah, M. A., Khader, A. T., & Abualigah, L. M. (2016). A comprehensive review: Krill Herd algorithm (KH) and its applications. Applied Soft Computing, 49, 437-446.
[31] Feng, Y., Wang, G.-G., Deb, S., Lu, M., & Zhao, X.-J. (2017). Solving 0–1 knapsack problem by a novel binary monarch butterfly optimization. Neural Computing and Applications, 28(7), 1619-1634.
[32] Wang, G.-G., Deb, S., Gao, X.-Z., & Coelho, L. D. S. (2016). A new metaheuristic optimisation algorithm motivated by elephant herding behaviour. International Journal of Bio-Inspired Computation, 8(6), 394-409.
[33] Wang, G.-G., Gandomi, A. H., Alavi, A. H., & Gong, D. (2019). A comprehensive review of krill herd algorithm: variants, hybrids and applications. Artificial Intelligence Review, 51(1), 119-148.
[34] Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019.
[35] Rohani, N., & Eslahchi, C. (2019). Drug-Drug interaction predicting by neural network Using integrated Similarity. Scientific reports, 9(1), 1-11.



Abdulsamad A. Al-Marghilani completed his M.S. degree in 2005 and Ph.D. in 2009 in Computer Science from De MontFort University, UK. He is currently working as an Associate Professor at Northen Border University, Arar in Saudi Arabia. His research interests include Artificial Intelligence, Neural Network, NLP, Text Mining, and Deep Learning. He has several publications in prestigious journals and conferences on the aforementioned topics.