Int. J. Nonlinear Anal. Appl.
Volume 12, Special Issue, Winter and Spring 2021, 2137-2151
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2021.6068

The optimal time slot selection and feature selection for the prediction of drugs for diseases

B. Nithya[a,*], G. Anitha[a]

[a]*Department of Computer Science, Karpagam Academy of Higher Education (Deemed to be University), Coimbatore-21, Tamilnadu, India*

(Communicated by Madjid Eshaghi Gordji)

Abstract

Data mining techniques have been applied to analyze, predict and diagnose diseases. The prediction of disease becomes meaningless when there is no proper recommendation of a drug to the patient. A drug recommendation method called Artificial Neural Network (ANN) with side effect constraints was proposed to recommend drug names for multiple diseases such as Chronic Kidney Disease (CKD), diabetic and heart disease based on the interaction between drug and disease and their side effects. In this drug recommendation method, multiple attributes of drugs and patients were collected from different sources and the hidden relationship between the attributes was predicted by using a Hidden Markov Model (HMM). In addition to this, statistical features were calculated and added as additional features. The collected and calculated features were used in ANN with side effect constraint classifier which predicted drug name for multiple diseases with the consideration of side effects. However, there is a high dimensionality problem in the recommended method due to more number of features. Moreover, it leads to more computational and space complexity in the ANN classifier. In this paper, an efficient Krill Herd (KH) algorithm for optimization is introduced to solve the above-mentioned problems in the drug recommendation method. According to the herding behavior of the likeness of the krill individuals, KH selects the optimal features. The multiple attributes of drugs and patients are collected in a different time slots. The KH algorithm is also used to select the optimal time slot. Then, the optimal time slot and features are given as input to ANN which predicts drug names for multiple diseases with high accuracy and low computational complexity.

Keywords: Drug recommendation, Feature selection, Optimization algorithm, Krill herd algorithm.

*Corresponding author
Email addresses: nithyabaalu87@gmail.com (B. Nithya), florenceanitha7@gmail.com (G. Anitha)

Received: October 2021 *Accepted:* December 2021

1. Introduction

Pharmaceutical science [14, 20] is an interdisciplinary domain of research incorporating different fields of science and engineering with an objective to discover a potential drugs. Discovery of Drug is a more complex process which involves in the identification of new drugs and their potential targets. Drug [7, 8, 21] can be considered as molecules that interact with an appropriate target protein to perturb biological systems of different molecular interactions such as signal transduction network, metabolic pathway and protein interaction network. Drugs are used to maintain and restore health by the prevention and treatment of illness. Generally, drugs are widely classified as brand-name and generic drug. Generic drugs are replications of brand-name drugs that have exactly the similar dosage, safety, cause of administration, risks and intended use as the original drug.

According to severity of disease, drugs are recommended to the patients. The potential side effects of the drugs are predictable from the chemical, biological and other related information. A method is required to recommend a drug to a patient with the consideration of side effect. A drug recommendation method called ANN with side effect constraint [12] was proposed to predict the drug name for multiple diseases such as CKD, diabetic and heart disease. In the drug recommendation method, different attributes of drugs and patients were collected. The hidden relationship among the features was calculated using a Hidden Markov Model (HMM). Then, the statistical measures were calculated and added as additional features. A side effect constraint was included with ANN classifier to predict the drug name with high accuracy. However, the computational complexity of the ANN classifier is high due to using more number of features in it.

So in this work, optimization algorithm Krill Herd (KH) [4] is introduced to select the most discriminative features from the collected attributes of drugs and patients such as drug name, combination (ingredients), size, shape, tablet coat, density, surface area, disintegration time, propensity for swelling, side effect, mode of drug (i.e., tablet, injection, capsules), cost of the medicines, drug class, dose, brand name, drug reaction, age of the patient, gender of the patient, weight of the patient and height of the patient. The attributes of drugs and patients are collected at different time slot which is also optimized by using the same KH algorithm. The KH algorithm consisted of a number of krill, their time-dependent position and it is formulated by the main three factors called movement induced by the presence of other individuals, foraging activity and random diffusion. The optimal time slot and the selected optimal features are processed in ANN with side effect constraint which predicts the drug name for multiple diseases. Thus, the proposed KH- ANN with side effect constraint reduces the computational and space complexity of ANN classifier.

2. Literature Survey

A prediction scheme [9] was proposed that combined Fuzzy preference based Rough Set (FPRS) method and semi-supervised SVM. The FPRS was utilized in gene markers from microarray gene expression to find out the cost relevant. FPRS reflected the degree of preference quantitatively making it more powerful in extracting information from fuzzy data than dominance or equivalence relation. This type of process was used to select optimal gene. Consequently, the selected gene markers were given as input to the Transductive Support Vector Machine (TSVM) which predicted cancer. TSVM was seeking the separation in long way in the presence of both labeled and unlabeled data through regularization process. Even though TSVM provided better performance by incorporating unlabeled data into the training set, the number of iterations is ambitious to define in TSVM.

A Sequential pattern mining [18] technique was proposed to predict the next prescribed medications. Initially, the data was collected from Blue Cross Blue Shield of Texas and it was categorized as training and testing set. A cSPADE was applied on the training data for identifying temporal

patterns of mediation prescribed for patients. The knowledge base of mined patterns was utilized for creating rules in which it predicts the next diabetes medication prescribed for a test set of patients. However, this technique is time consuming and expensive process.

A Scored Mean Kernel Fusion (SMKF) [10] was proposed which was the computational method used for prediction of drug-disease associations using fusion of information and aggregation. The dataset consisted of multiple features collected from diverse data sources which provides better accurate and reliable results. In the proposed computational method, feature fusion method was used to make high-level features. The SMKF method was utilized score mean which systematically combined multiple features which relates to diseases or drugs at two levels to predict the indication of drugs. The drug disease can be improved by using various other features of drugs such as the pathway feature.

A velocity bounded Boolean Particle Swarm Optimization (VbBoPSO) and Improved VbBoPSO (IVbBoPSO) [5] was proposed for improved feature selection in liver and kidney disease diagnosis. VbBoPSO was identical to BoPSO except that it is checked for velocity bound and keeps the length of each velocity component within the range of maximum velocity and minimum velocity. This kind of algorithm may lead to dullness during subsequent iterations. This problem was solved by IVbBoPSO where the new particles were introduced if the particle's global best position or particle's best solution was not updated for a certain number of repetitive iterations. The inspection was increased and there was an opportunity of getting an optimal solution by searching in the new search space. However, a parameter used for good convergence introduced had less complications of tuning of a parameter.

A new method of feature selection [15] was proposed based on Multi-Objective Particle Swarm Optimization (MOPSO) and Non-dominated Sorting Genetic Algorithm-II (NSGGA-II) along with Artificial Neural Network (ANN) for dose prediction. In NSGGA-II, all members of genetic population has been compared with each other and had been placed within the categories. First category members that do not dominate by other members of the population. The second members of the category were beaten by members of the first category and this process was iterated in the same way in the other categories to rank all members in the category. The MOPSO algorithm selected the optimal features based on grid index. The concurrence speed of this algorithm was great. The selected features were used in ANN for dose prediction. However, the MOPSO has fewer errors. A link prediction approach [6] was proposed for the drug recommendation in a disease-drug bipartite graph. A threshold value was used in this approach to assure that the crucial links were not lost and unwanted links were not taken into considerations. However, this threshold value greatly influences the performance of link prediction approach.

For prediction of drug-disease interactions, a semi-supervised graph cut algorithm and three-layer data integration [**?**] method was proposed. However, similarity integration strategy was used in the method that has an unsteady point that neglects the information from the layer with much smaller sores. Bipartite Local Model (BLM) with a hubness-aware regression technique [2] was proposed for drug-target interactions. The features in the collected data were extracted and based on these features a projection based ensemble of BLMs was constructed which predicts the drug-target interactions. However, it has low precision. Bayesian Ranking Prediction of Drug-target Interactions (BRDTI) [12] was proposed for drug-target prediction. However, this method is for under improvement as few of the new interactions was neglected by some estimation method.

A new hybrid approach [1] was proposed to select the most discriminative features in the dataset. Here, a framework was presented with two components. Self-Adaptive Cohort Intelligence (SACI) algorithm was proposed and which was the first component of the framework. It was the modified version of Cohort Intelligence (CI) meta-heuristic algorithm. Self adaptive scheme was utilized in

SACI algorithm for computation of sampling rate and the tournament-based mutation was used for computation of mutation rate. Then SACI was integrated with Support Vector Machine (SVM) and named as SACI- SVM was used for feature selection. However, search strategies based on SACI-SVM leads to high computation time for high- dimensional datasets.

Machine learning based drug indication prediction [3] method was proposed using Linked open data. A binary feature matrix was created using drug target, substructure and side effects and disease ontology terms. Various classifiers such as Linear Regression (LR), K-Nearest Neighbor (KNN), Random Forest (RF) and Gradient Boosting Classifier (GBC) were classifiers were processed for drug indication prediction. However, the efficiency of this method is low.

A modified Differential Evolution (DE) algorithm [17] was proposed for selection of feature in a prediction of heart disease. It selects the critical features that forms the foundation cause for the objective function of the heart disease prediction problem. The conventional DE randomly selected the individual vectors which have weight value and then it finds the difference between the first two vectors which was added to the third vector which determines the mutant vector. In the modified DE algorithm, four vectors were sequentially selected and then computes the weighted difference between the first two vectors and the next two vectors in a parallel manner. The obtained two weighted difference vectors were added to generate the mutant vector. However, an error is present in the prediction of heart disease.

A novel feature selection strategy called Genetic Algorithm (GA) [16] was proposed for dimensionality reduction. The genetic programming was volatile to skewness of data issues and it works well with both of the balanced and unbalanced data. The proposed strategy combines the most discerning features selected by distinct feature selection metrics which generates different feature space projection based on different criteria it was influenced by several factors such as data skewness. Instead of combined metrics in an exhaustive way, the proposed strategy combines the generated projections on sparse and highly dimensional skewed data. However, the efficiency of this strategy is low.

A variant of firefly algorithm [20] was proposed to select the most important features which were used in classification and regression models. The problem in original firefly algorithm was getting into local optima and premature convergence problem. This was solved by the proposed variant of firefly algorithm which employs Simulated Annealing (SA). It enhances global and local promising solutions chaotic-accelerated attractiveness parameters and diversion mechanisms of weak solution. An application programming interface [11] was provided for recommending drugs to the users suffering from a particular disease. The inescapable scaling of clinical information builds the potential for information mining systems which enhances the quality and reduces the cost of social insurance.

3. Proposed Methodology

In this fragment, the proposed methodology for selection of feature and time slot optimization in a prediction of drug name for multiple diseases is described in detail. From the various sources, different data about patients and drugs such as drug name, combination (ingredients), size, shape, tablet coat, density, surface area, disintegration time, propensity for swelling, side effect, mode of drug (i.e., tablet, injection, capsules), cost of the medicines, drug class, dose, brand name, drug reaction, age of the patient, gender of the patient, weight of the patient and height of the patient are collected. These data are collected at different time slots. For example, the data can be collected for every 3 days or 5 days or 7 days of time interval. A Krill Herd (KH) algorithm is introduced

which simultaneously selects both the optimal features and time slot which are used in ANN with side effect constraint classifier to predict the drug name for multiple diseases.

3.1. Krill Herd Based Feature Selection and Time Slot Optimization

The new swarm intelligent optimization algorithms is Krill Herd (KH) optimization algorithm. This algorithm imitates the performance of krill where each individual in the krill herd will make its own contribution in the moving process depending upon the fitness. Moreover, it mainly depends on whether adjacent krill individual possesses this repulsive or attractive effect on each individual which act as a local search for each. The food center is identified from the overall fitness of the krill individual and that is considered as the best global estimation. Based on the krill individual movement induction, foraging activity and random diffusion, the time- dependent location of an individual krill is calculated.

A d-dimensional search space Lagrangian model has been adopted by the KH algorithm.

$$\frac{dX_i}{dt} = M_i + F_i + R_i \tag{3.1}$$

where M_i refers to the krill individual movement induction, F_i denotes the foraging activity and R_i denotes the random diffusion.

The KH algorithm is started with initializing different parameters such as maximum number of iteration $\max_{Itr}$, maximum induced speed $\max_m$, maximum diffusion speed $\max_R$, maximum foraging speed $\max_F$, number of krill NUM, position of krill X and time slot T and number of features NUM_F. The initialized number of features and time slot is equal to the number of krill. After the initialization process, KH algorithm generates the location of features and time slot randomly. For each position of current features and time slot, the fitness (ANN classification accuracy) of each krill is calculated.

3.1.1. Krill Individual Movement Induction

The communicative effects within the krill individuals lead to the movement for as much as they always experiment to preserve a high density. The progress due to the other krill individuals is calculated by using the following equation:

$$M_i^{new} = \max_M \delta_i + W_{num} M_i^{old} \tag{3.2}$$

where W_{num} represents the inertia weight with the motion induced and has a value in the range from 0 to 1 and M_i^{old} denotes the last motion induced.

$$\delta_i = \delta_i^{local} + \delta_i^{target} \tag{3.3}$$

where δ_i^{local} is the neighbor's local effect which is calculated as,

$$\delta_i^{local} = \sum_{j=1}^{Num_n} \hat{K}_{i,j} \hat{X}_{i,j} \tag{3.4}$$

where,

$$\hat{X}_{i,j} = \frac{X_j - X_i}{||X_j - X_i|| + t} \tag{3.5}$$

$$\hat{K}_{i,j} = \frac{K_i - K_j}{K_{worst} - K_{best}} \tag{3.6}$$

where K_{best} and K_{worst} are the best and worst krill individual values respectively, K_i is the fitness (ANN accuracy) of i-th individual krill, K_j is the fitness of j-th neighbor, t is a small positive number and Num_n denotes the number of neighbors.

The effect of target direction δ_i^{target} is handed over by the best krill individual and it is calculated as follows:

$$\delta_i^{target} = A_{best}\hat{K}_{i,best}\hat{X}_{i,best} \tag{3.7}$$

where A_{best} denotes the effective coefficient of krill individual with best fitness to i-th individual krill which is determined as follows:

$$A_{best} = 2\left(r + \frac{Itr}{Itr_{\max}}\right) \tag{3.8}$$

where, r is the random value between 0 to 1 and Itr denotes the actual iteration number. The sensing distance is calculated as follows:

$$dist_{s,i} = \frac{1}{5N}\sum_{j=1}^{N} X_i - X_j \tag{3.9}$$

where N denotes the no of krill individuals and $X_i(X_j)$ denotes the correspondent position of the i(j)-th krill. If the distance between X_i and X_j is lesser than the specified sensing distance then X_j is a neighbor of X_i.

3.1.2. The Movement Due to Foraging Activity

The foraging motion is defined based on the location of the food as well as the previous experience about the location of the food. The foraging motion is given as follows:

$$F_i = \max_F \varepsilon_i + w_F F_i^{old} \tag{3.10}$$

where w_F denotes the inertia weight of foraging motion.

$$\varepsilon_i = \varepsilon_i^{food} + \varepsilon_i^{best} \tag{3.11}$$

where ε_i^{best} denotes i-th krill's best fitness so far and ε_i^{food} denotes the food attractive which is given by,

$$\varepsilon_i^{food} = A^{food}\hat{K}_{i,food}\hat{X}_{i,food} \tag{3.12}$$

where A^{food} is the food coefficient which is calculated as:

$$A^{food} = 2\left(1 - \frac{Itr}{Itr_{\max}}\right) \tag{3.13}$$

$$\varepsilon_i^{best} = \hat{K}_{i,ibest}\hat{X}_{i,ibest} \tag{3.14}$$

where $\hat{K}_{i,ibest}$ and $\hat{X}_{i,ibest}$ denotes the i-th value of previously visited best fitness and position of krill individual's respectively.

Finally in the process of migration due to the foraging activity, the food center is calculated for each iteration as follows:

$$X^{food} = \frac{\sum_{i=1}^{N}\frac{1}{K_i}X_i}{\sum_{i=1}^{N}\frac{1}{K_i}} \tag{3.15}$$

3.1.3. The Physical Diffusion

The Physical diffusion motion is determined in the terms of diffusion speed at the maximum and a random directional vector which is given as follows:

$$R_i = \max_R \left(1 - \frac{Itr}{Itr_{\max}}\right) \beta \tag{3.16}$$

where β denotes the random directional vector.

3.1.4. Updating Position

The positions of krill individuals are updated for each time slot and a different number of features. Moreover, the performance of KH algorithm is improved by integrating the mechanisms of genetic reproduction called the crossover and mutation with KH algorithm.

Crossover

$x_{i,s}$ is the s-th component of X_i which is calculated by using the following equation:

$$x_{i,s} = \begin{cases} x_{t,m}, r_{i,m} < C_0 \\ x_{i,m},\ else \end{cases} \tag{3.17}$$

where, C_0 is the crossover probability which is equal to 0.2 $\hat{K}_{i,ibest}$.

Mutation

$$x_{i,m} = \begin{cases} X_{gbest,m} + > \rho(x_{p,m} - x_{q,m}), r_{i,m} < M \\ x_{i,m},\ else \end{cases} \tag{3.18}$$

where, M denotes the mutation probability which is equal to $\frac{0.05}{\hat{K}_{i,ibest}}$.A krill individual's position vector during the interval $[tm, tm + \Delta tm]$ is calculated as follows and which is based on the different parameters of the movement:

$$X_i(tm + \Delta tm) = X_i(tm) + \Delta tm \frac{dX_i}{dt} \tag{3.19}$$

This process is continued until the no of iterations reaches the maximum. Finally, the optimal features and time slot are selected and it is associated with the total best krill (time slot and features). The optimized time slot and features are used in ANN with side effect constraint algorithm which predicts the drug name for CKD, diabetic and heart disease. By using optimal features and time slot, the computational and space complexity of the classifier is reduced as well as the accuracy of the drug recommendation method is increased.

Algorithm

Step 1: Collect drug and patient features

Step 2: Find the hidden relation among features and calculate statistical measures.

Step 3: Initialize $\max_{Itr}$, $\max_M$, $\max_R$, $\max_F$, Num, X and Num_F

//Position Calculation

Step 4: Generate the position of features and time slot randomly.

//Fitness Calculation

Step 5: Calculate the fitness (ANN with side effect constraint accuracy) of the current time slot and features position.

Table 1: Brand-Name Drugs for CKD

Medicine Name	Class Name	Dosage	Side Effects	Ingredients
Edarbi	ARBs	80 mg	nausea, diarrhea, fatigue, cough, dizziness on standing	sodium hydroxide, microcrystalline cellulose, mannitol, fumaric acid, croscarmellose sodium, hydroxypropyl cellulose, and magnesium stearate
Cozaar	ARBs	50 mg	dry cough, muscle cramps, tired feeling	microcrystalline cellulose, lactose hydrous, pregelatinized starch, magnesium stearate, hydroxypropyl cellulose, hypromellose, and titanium dioxide
Teveten	ARBs	600 mg	headache, dizziness, tired feeling	crospovidone, hypromellose, lactose monohydrate, magnesium stearate, microcrystalline cellulose, polyethylene glycol, polysorbate 80, pregelatinized starch, and titanium dioxide

//Motion Calculation

Step 6: Calculate the motion which is induced by other individual using equations from (3.2) to (3.9).

Step 7: Compute foraging motion using equations from (3.10) to (3.15).

Step 8: Evaluate physical diffusion motion using equation (3.16).

Step 9: Update positions for each time slot and features using equations from (3.17) to (3.19).

Step 10: while ($Itr == \max_{Itr}$)

Step 11: Select the time slot and features associated with the overall best Krill.

4. Results and Discussions

In this section, the effectiveness of the proposed drug prediction method is evaluated in the terms of accuracy, precision, and the recall. The effectiveness of existing and proposed drug prediction method is tested using Java 8. For the experimental purpose, different information of patients and drugs are collected from different online resources, hospitals etc. The sample data are shown in Table 1,2,3,4,5,6.

4.1. Accuracy

Accuracy tends to the measure of correctly selects the optimal time slot and features to recommend a drug name in all instances. It can be calculated by,

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

Table 7 tabulates the accuracy values of proposed KH-ANN with side effect constraint and existing ANN with side effect constraint.

Table 2: Generic Name Drugs for CKD

Medicine Name	Class Name	Dosage	Side Effects	Ingredients
Enalapril	ACE-I	5-40 mg	sleep problems (insomnia), dry mouth, nausea, vomiting, diarrhea	lactose, magnesium stearate, sodium bicarbonate, and starch. The 10 mg and 20 mg tablets also contain iron oxides
Irbesartan	ARBs	150-300 mg	drowsiness, confusion, mood changes, increased thirst, loss of appetite, nausea and vomiting, swelling, weight gain, feeling short of breath	lactose, microcrystalline cellulose, pregelatinized starch, croscarmellose sodium, poloxamer 188, silicon dioxide, and magnesium stearat
Lisinopril	ACE-I	10-40 mg	Cough, dizziness, drowsiness, headache, depressed mood	calcium phosphate, mannitol, magnesium stearate, and starch. The 10 mg and 20 mg tablets also contain iron oxide

Table 3: Generic Drugs for Diabetic Disease

Medicine Name	Class Name	Dosage	Side Effects	Ingredients
Tolbutamide	Sulfonylureas	500 mg	Hypoglycemia, weight gain	Colloidal silicon dioxide, magnesium stearate, microcrystalline cellulose, sodium lauryl sulfate and sodium starch glycolate.
Gilmepiride	Sulfonylureas	1 mg, 2 mg, 4 mg	Hypoglycemia, weight gain	Lactose (hydrous), sodium starch glycolate, povidone, microcrystalline cellulose, and magnesium stearate.
Glipizide	Sulfonylureas	5 mg, 10 mg	Hypoglycemia, weight gain	Lactose, microcrystalline cellulose, starch and stearic acid. Minodiab 5 mg tablets are white, biconvex tablets scored on both sides.

Table 4: Brand-Name Drugs for Diabetic Disease

Medicine Name	Class Name	Dosage	Side Effects	Ingredients
glyset	alpha-glucosidase inhibitors	25 mg	Abdominal pain, diarrhea, and flatulence	Starch, microcrystalline cellulose, magnesium stearate, Hypromellose, polyethylene glycol, titanium dioxide, and polysorbate 80.
Gliclazide	Sulphonylureas	80 mg	Hypoglycemia (low blood sugar): Gliclazide, like other sulfonylurea drugs, can cause symptoms of hypoglycemia (low blood sugar) including dizziness, lack of energy, drowsiness, headache, and sweating have been observed. Weakness, nervousness, shakiness, and numbness or tingling	Each tablet contains 80mg of gliclazide. The other ingredients are maize starch, stearic acid, magnesium stearate, microcrystalline cellulose (E460).
Glimepiride	Sulphonylureas	1 mg, 2 mg and 4 mg	• severe skin rash, itching, redness, or irritation • pale skin, easy bruising or bleeding, fever, unusual weakness • numbness or tingly feeling • trouble breathing • feeling like you might pass out • dark urine, clay-colored stools	lactose (hydrous), sodium starch glycolate, povidone, microcrystalline cellulose, and magnesium stearate

Table 5: Brand-Name Drugs for Heart Disease

Medicine Name	Class Name	Dosage	Side Effects	Ingredients
Ikorel	Potassium Channel Openers	10 mg to 20 mg twice a day	Headache, dizziness and light-headedness, especially when getting up from a sitting or lying down position, tiredness, drowsiness, cough	maize starch, croscarmellose sodium, stearic acid, mannitol
Efient	Antiplatelet Drugs	60-mg	an increased tendency for bleeding, headache, dizziness, back pain	mannitol, Hypromellose, low-substituted hydroxypropyl cellulose, microcrystalline cellulose, sucrose stearate, and glyceryl behenate
Innovace	ACE Inhibitors	2.5 mg	swelling of your hands, feet or ankles, cough	lactose monohydrate, sodium hydrogen carbonate, maize starch, pregelatinized starch, magnesium stearate

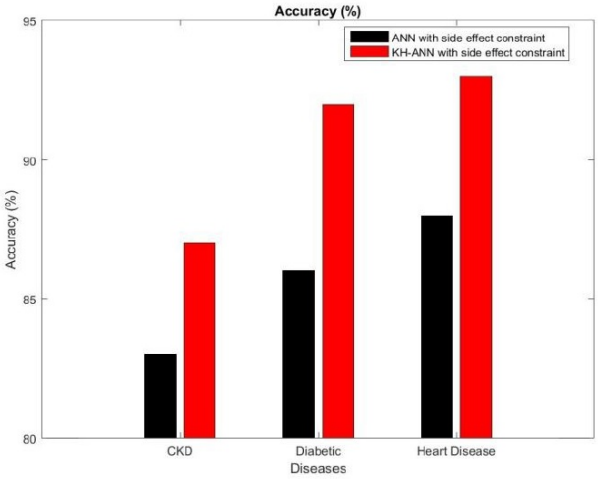

Figure 1: Comparison of Accuracy (%)

Table 6: Generic Drugs for Heart Disease

Medicine Name	**Class Name**	**Dosage**	**Side Effects**	**Ingredients**
doxazosin (dox AY zo sin)	alpha blockers	8mg	fast and pounding heartbeat, irregular heartbeat, Dizziness	microcrystalline cellulose, lactose, sodium starch glycolate, magnesium stearate and sodium lauryl sulfate
digoxin	Cardiac glycosides	10 to 15 mcg/kg	slow heartbeat, dizziness, fainting	Each 2-mL ampule of LANOXIN Injection contains 500 mcg (0.5 mg) digoxin (250 mcg [0.25 mg] per mL)
Furosemide	Diuretics	20 to 80mg	weight loss, body aches, fever, wheezing, nausea and vomiting	lactose monohydrate NF, magnesium stearate NF, starch NF, talc USP, and colloidal silicon dioxide NF

Table 7: Comparison of Accuracy (%)

Diseases	**ANN with side effect constraint**	**KH-ANN with side effect constraint**
CKD	83	87
Diabetic	86	92
Heart disease	88	93

Table 8: Comparison of Precision

Diseases	ANN with side effect constraint	KH-ANN with side effect constraint
CKD	0.811	0.854
Diabetic	0.846	0.887
Heart disease	0.839	0.875

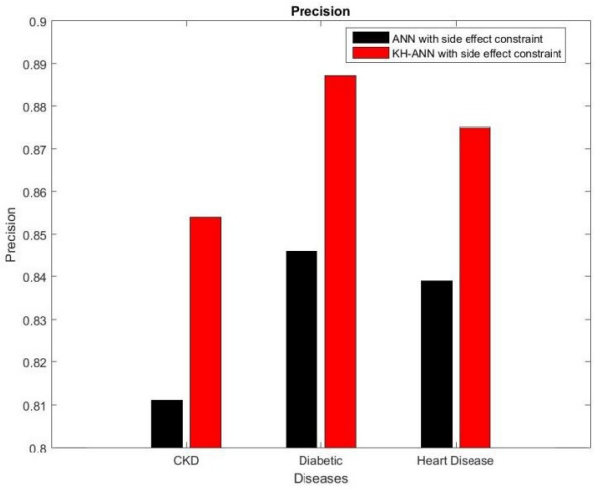

Figure 2: Comparison of Precision

Figure 1, comparison of KH-ANN with side effect constraint and ANN with side effect constraint in terms of accuracy. It is clearly known from the graph is that the proposed KH- ANN with side effect constraint has high accuracy than the ANN with side effect constraint for three different types of diseases are CKD, diabetic and heart disease. Clearly, it proves that the optimized time slot and features increases the accuracy of drug name prediction method.

4.2. Precision

The Precision value is calculated bestow to the relevant information at true positive prediction value and false positive prediction value.

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

Table 8 tabulates the precision values of proposed KH-ANN with side effect constraint and existing ANN with side effect constraint.

Figure 2, comparison of KH-ANN with side effect constraint and ANN with side effect constraint in terms of precision. It is clearly known from the graph is that the proposed KH- ANN with side effect constraint has high precision than the ANN with side effect constraint for three different types of diseases are CKD, diabetic and heart disease. Clearly, it proves that the optimized time slot and features increases the precision of drug name prediction method.

4.3. Recall

The value of recall is calculated bestow to recommend the drug name at true positive value and false negative value predictions.

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

Table 9: Comparison of Recall

Diseases	ANN with side effect constraint	KH-ANN with side effect constraint
CKD	0.80	0.84
Diabetic	0.84	0.89
Heart disease	0.82	0.87

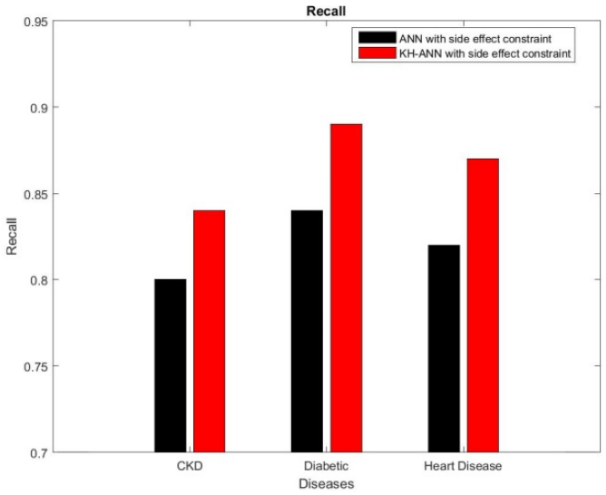

Figure 3: Comparison of Recall

Table 9 shows the recall values of proposed ANN with side effect constraint and existing ANN without side effect constraints.

Figure 3, comparison of KH-ANN with side effect constraint and ANN with side effect constraint in terms of recall. It is clearly known from the graph is that the proposed KH- ANN with side effect constraint has high recall than the ANN with side effect constraint for three different types of diseases are CKD, diabetic and heart disease. Clearly, it proves that the optimized time slot and features increases the recall of drug name prediction method.

5. Conclusion

In this paper, a drug recommendation method is improved by introducing a Krill Herd (KH) optimization algorithm. Initially, data related to drug and patients are collected from different sources. The hidden relationship among features and statistical measures are calculated and added as additional features. The KH algorithm selects the optimal time slot and features which is based on the simulation of behavior of herding of the krill individuals. It searches for which time slot and for which features, the classification accuracy of Artificial Neural Network (ANN) is high. The selected time slot and features by KH algorithm are used in ANN to predict the drug name for CKD, diabetic and heart disease. The experimental result shows that the proposed KH- ANN with side effect constraint has high accuracy, precision and recall for prediction of drug name for CKD, diabetic and heart disease.

References

[1] M. Aladeemy, S. Tutun and M.T. Khasawneh, *A new hybrid approach for feature selection and support vector machine model selection based on self-adaptive cohort intelligence*, Expert Systems with Applications, 88 (2017) 118–131.

[2] K. Buzaand and L. Peska, *Drug–target interaction prediction with Bipartite Local Models and hubness-aware regression*, Neurocomput. 260 (2017) 284–293.
[3] R. Celebi, O. Erten and M. Dumontier, *Machine Learning Based Drug Indication Prediction Using Linked Open Data*, SWAT4LS, 2017.
[4] A.H. Gandomi and A.H. Alavi, *Krill herd: a new bio-inspired optimization algorithm*, Commun. Nonlinear Sci. Numerical Simul. 17(12) (2012) 4831–484.
[5] S. Gunasundari, S. Janakiraman and S. Meenambal, *Velocity bounded boolean particle swarm optimization for improved feature selection in liver and kidney disease diagnosis*, Expert Systems with Applications, 56 (2016) 28–47.
[6] E. Gundogan and B. Kaya, *A link prediction approach for drug recommendation in disease-drug bipartite network*, Artific. Intell. Data Process. Symp. 2017 pp. 1–4.
[7] M.J. Jahid and J. Ruan, *An ensemble approach for drug side effect prediction, bioinformatics and biomedicine (BIBM)*, 2013 IEEE Int. Conf. IEEE (2013) 440–445.
[8] M.H.M. Lau and J.W. Tenney, *Evaluation of drug-disease interactions and their association with unplanned hospital readmission utilizing STOPP version 2 criteria*, Geriatrics 2 (2017) 33.
[9] U. Maulik and D. Chakraborty, *Fuzzy preference based feature selection and semisupervised SVM for cancer classification*, IEEE transactions on nanobioscience, 13(2) (2014) 152–160.
[10] H. Moghadam, M. Rahgozar and S. Gharaghani, *Scoring multiple features to predict drug disease associations using information fusion and aggregation*, SAR and QSAR in Environmental Research, 27(8) (2016) 609–628.
[11] V. Mudaliar, P. Savaridaasan and S. Garg, *Disease prediction and drug recommendation android application using data mining (virtual doctor)*, Int. J. Recent Technol. Engin. 8(3) (2019) 6996–7001.
[12] L. Peska, K. Buza and J. Koller, *Drug-target interaction prediction: A bayesian ranking approach, Computer methods and programs in biomedicine*, 152 (2017) 15–21.
[13] A. Sharma and R. Rani, *BE-DTI': Ensemble framework for drug target interaction prediction using dimensionality reduction and active learning*, Computer Methods and Programs in Biomedicine, 165 (2018) 151–162.
[14] M.K. Sohrabi and A. Tajik, *Multi-objective feature selection for warfarin dose prediction*, Computational biology and chemistry, 69 (2017) 126–133.
[15] F. Viegas, L. Rocha, M. Gonçalves, F. Mourão, G. Sá, T. Salles, G. Andrade and I. Sandin, *A genetic programming approach for feature selection in highly dimensional skewed data*, Neurocomputing, 273 (2018) 554–569.
[16] T. Vivekanandan and N.C.S.N. Iyengar, *Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease*, Computers in biology and medicine, 90 (2017) 125–136.
[17] A.P. Wright, A.T. Wright, A.B. McCoy and D.F. Sittig, *The use of sequential pattern mining to predict next prescribed medications*, J. Biomed. Inf. 53 (2015) 73–80.
[18] G. Wu, J. Liu and C. Wang, *Predicting drug-disease interactions by semi-supervised graph cut algorithm and three-layer data integration*, BMC Med. Genom. 10(5) (2017) 79.
[19] Y. Yang, Z. Ye, Y. Su, Q. Zhao, X. Li and D. Ouyang, *Deep learning for in vitro prediction of pharmaceutical formulations*, Actapharmaceuticasinica B 9(1) (2019) 177–185.
[20] L. Zhang, K. Mistry, C.P. Lim and S.C. Neoh, *Feature selection using firefly optimization for classification and regression models*, Decision Support Systems, 106 (2018) 64–85.
[21] D. Zhou, L. Miao and Y. He, *Position-aware deep multi-task learning for drug–drug interaction extraction*, Artific. Intell. Med. 87 (2018) 1–8.