Radial Subset Clustering Feature-Based Deep Spectral Neural Classification for relational drug recommendation

S. Dinakaran[1*], V. Ravi[2], P. Anitha[3]

[1]Research Scholar, K.S.R College of Engineering, Namakkal, Tamilnadu, India.
[2]Professor, Department of Electrical and Electronics Engineering, K.S.R College of Engineering, Namakkal, Tamilnadu, India.
[3]Professor, Department of MCA, K.S.R College of Engineering, Namakkal, Tamilnadu, India.
E-mail: [1]ksrdinati201@yahoo.com, **E-mail:** [2]ervravi@gmail.com, **E-mail:** [3]psp03ster@gmail.com

Abstract.

Paramedical science is the developing strategies in the medical field for growing drug patterns to create new compound molecules. The compound molecules are non-relation to suggest the drug to the patients to cause side effects. This accuses of compound molecule features being non-depended with each other doesn't create patterns to classification inaccuracy. To resolve this problem, we propose a relative drug compound analysis based on Radial Subset Clustering Feature-Based Deep Spectral Neural Classification (RSCF-DSNC) for predicting relational drug recommendation. Initially, the dataset is collected from paramedical compound dataset related to paracetamol, and diphenhydramine hydrochloride bioinformatics data are pre-processed. Then the featured labels are marginalized to estimate the Relative Drug Intensive Weight (RDIW) for finding the mutual correlation. Based on the correlation weight, the Radial Subset Clustering Feature Selection (RSCFS) is applied to predict the relative closeness of drug molecules. Infra Segment Subset Feature Pattern Theory (ISSFPT) technique is used to analyze the successive pattern weight from the RSCFS dataset. Then the selective feature patterns are then trained into Deep Spectral Neural Classification (DSNC) adapted with Convolution Neural Network (CNN) to identify the relational class of drug molecules category. The proposed drug success rate prediction performance result is 93% low time complexity 31sec. This produces high prediction accuracy compared to the other system than other methods for recommending features for relation to drug compound recommendation.

Keywords: Drug compound analysis, pattern prediction, subset clustering, feature selection

DOI Number: 10.14704/nq.2022.20.9.NQ440097 **Neuro Quantology 2022; 20(9):873-883**

1. Introduction

A drug recommendation system is a similar system that recommends drugs for a specific illness based on patient reviews [1]. This system is essential in rapidly evolving technology and can save lives by helping doctors. Side effects of a drug are the human reaction to conventional doses of dangerous and unwanted drugs used to prevent, diagnose, treat, and alter the physiological function of the disease.

The recommended drug in medical productiveness is a non-residual form of inaccurate compound analysis, which depends on many reasons that cause faults in human health. It accuses of compound molecule features being non-depended with each other doesn't create patterns to classification inaccuracy. To address this problem, data analysis based on drug compound recommendations is required. These recommendations, along with quality data, find compilations of high-level Deep Learning (DL) based related features that can improve a given molecule's innovation and decision-making.

The proposed algorithm aims to identify the recommendation of the effective drug using the deep learning technique. Diseases caused by inappropriate drug compound molecules can cause by receptors, enzymes, hormone receptors, or other functional proteins and functional alterations.

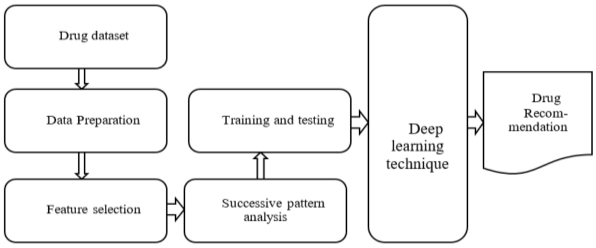

Figure 1: Predicting relational drug recommendation work process

The proposed pre-processing is done to remove noise and check null values from the given dataset. The pre-processed dataset fed into Relative drug intensive weight calculates, Radial Subset Clustering Feature Selection (RSCFS) to select the relevant features. Then creating successive patterns using ISSFPT. Finally, a Deep Spectral Neural Classification (DSNC) with CNN algorithm is done to classify the drug recommendation shown in figure 1. Selects the best one for drug recommendation system to improve drug prediction accuracy, precision, and model efficiency. The proposed algorithm efficiently solves the above problems and provide high performance.

2. Related work

A. Ezzat, P (2017) focuses on reviews of drug repositioning using various approaches such as network-based cluster, propagation, Text Mining (TM) semantic-based approaches for paramedical science. Similarly, the author proposes Collaborative Metric Learning (CML) algorithm to capture the significant relationship between the drug object it used to the recommendation system for paramedical science A.Wen,(2020). Nonetheless, these approaches can't provide effective recommendation results.

D. Afdhal (2020) presents Chi-Square Mutual Information (CSMI) and Information Gain (IG) method for selecting the best features of drug name recognition. Similarly, the author focuses on selecting drug features using the Fuzzy Rough Feature Selection (FRFS) method to make correct decisions for drug segmentation. F. Ali et al (2017) focused the Data-Driven Feature Selection (DDFS) approach was used for drug density prediction. But these methods are only focused on feature selection of drug recommendation systems, so it does not provide classification result G. Jin (2014).

H. Luo (2019) focuses on drug selection in medical practice using a Fuzzy System (FS) based on medical prescription data. H. Xue, (2018) presents in Support Vector Machine (SVM) for social features. SVM was used to extract the essential elements of diabetes drugs.

J. Yang (2014) focus on Drug-Repositioning Pipelines (DRP) for taking turns application using the elucidated mechanism. J. Y. Ryu (2018) presents in Long Short-Term Memory Network (LSTM) was used to extract the relevant information based on Drug-Drug Interaction (DDI) and classify the recommended drugs.

Koras, K (2020) presents the Multi-label Linear Discriminant Analysis (MLDA) and Multi-Label Learning (MLL) algorithm for drug reaction prediction. The author focuses on an interactive computational diverse pipeline named DTINet for drug recommendations.

M. Fedorová (2018) introduces the machine learning-based approach such as K-Means Clustering (KMC), Random Forest Classification (RFC), and Multi-Layer Perceptron Classification (MLPC) model to analyze the drug expiration using the Titanic dataset. But in that study doesn't provide accurate classification results. M. L. Shahreza (2017) proposes a low-Rank Matrix Approximation (LRMA) and Drug Repositioning Recommendation System (DRRS) approach. LRMA was used to remove unknown entries and check null values. DRRS method was used to evaluate the information of drug and disease.

M. Tsubaki (2019) suggests a Heterogeneous Network Model (HNM) approach for drug repositioning, and it incorporates drug target information. Likewise, N.J. Perualila (2019) Heterogeneous Label Propagation (HLP) algorithm predicts the target disease-drug information.

Shengyu Liu (2025) presents in the Causal Inference-Probabilistic Matrix Factorization (CIPMF) method to classify the disease-drug and repositioning. Similarly, SwathikClarancia Peter (2019) introduces the Graph Regularized Matrix Factorization (GRMF) to predict the target disease drug recommendation. But these methods have a more difficult task during recommendation systems.

T. Chen (2019) introduces the deep DDI method used to correctly predict the DDI types to a given drug pair. Similarly, T. He (2017) defined as, GESSE algorithm to analyze and predict the drug side effects. T. Pahikkala (2015) introduces in Machine learning techniques and the MLL method for predicting drug reactions. But these methods have more challenging tasks, and it takes much time to provide a drug recommendation system.

3. Materials and methods

The proposed drug compound analysis based on the Radial subset clustering feature using deep spectral neural classification screening in drug innovation increases the data composition data due to the rapid growth of high-performance screening. This combination is one of the critical technologies in the drug compound analysis that require a feature selection method that requires a paramedical data set. The proposed system predicts accurate drug recommendation using Deep Spectral Neural Classification (DSNC) algorithm. One of the most used compound selection methods is the Radial Subset Clustering Feature (RSCF) which includes a set of compounds as clusters and choosing a few compounds from each group.

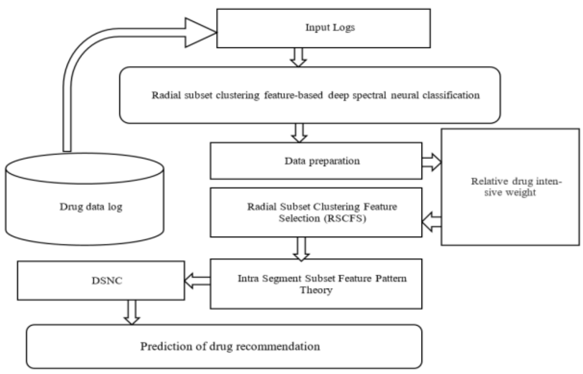

Figure 2: Architecture diagram for proposed RSCF-DSNC

Figure 2 describes the proposed diagram of drug recommendation, Radial subset clustering feature-based deep spectral neural classification analysis for a recommendation. The data preparation and drug weight are calculated by the RSCFS, In the ISSFPT is feature pattern recognizing and then DSNC for attained the optimizing results.

This classification creates the best recommendations with integrated neurological networks, which creates settings that make up the settings that can find high-level functions with integrated analysis. The importance of deep neuroscience using a cluster of the feature is dependent on the dimension and complexity of the compressed network during the choice of quality. The logical suggestion is trained to find the closest feature of the logical weight, and the expected release difference is evaluated logically. Each release is similar to a task (or class) that should predict. If the release layers are only one node, it is a classified class label to produce a drug class.

3.1 Data preparation

Data preparation is the process of removing inaccurate, incomplete, and inaccurate data from the dataset and replacing the missing values. In this phase, the first process for data preparation is done to noise removal and checking null values.

Algorithm steps for data preparation

Input: Drug Compound dataset (Sa) and combinatorial molecule (StBd)

Output: pre-processed dataset (PDS)

Begin

Step 1: Initialize the Collective dataset 'sa' and 'StBd'

Step 2: Read all dataset records Rd➔Sa1, Sa2, Sa3… + StBd1, StBd2… .

Create Index IistDx→Sa+StBd(Index+1)

Step 3: For (Rd combines attributes as feature Fi→Dx)

Check →Null field if Yes Terminate record and update recordset Rcs.

Else

Return Rcs←Sa+StBd;

End if;

End for;

Step 4: Compute Duplicate and non-fill case Rcs

Remove duplicate if Rcs→ Repeat row(Rw)

Update Rcs

End if

Check if (Rcs!=Empty fill near average)

Step 5: Get all the future index Rcs.

Return Fdx= $F(Sa + StBd)^{Fi} = \sum_{I=0}^{n} \binom{n}{k} Rcs.Update$→ gets real index, where n is total dominant features with k observed feature.

Return redundant feature set-Rds←Rcs(Fdx)

End if

Stop

The above Pseudo code prepares to process the original data for noise reduction and the collective student dataset from bipolar disorder and academic performance as features that are ordering the feature index with redundant features. Let assume Rcs presents dataset row and column set, Fi refers to features.

3.2 Relative Drug Intensive Weight (RDIW)

Relative Drug Intensive Weight (RDIW) is used for feature divides a set of time series data into several subgroups so that each Subset is compiled individually. During modular processing of the entire dataset, the dominant clusters within each Subset collect. The newly created reduced database contains patterns from the Subset cluster. The final Relative Drug Intensive Weight (RDIW) is performed in classification from the Subset to identify the desired input-output data.

Algorithm steps

Input: Pre-processed dataset (PDS)

Output: Drug-intensive weightD_w

Begin

Step 1: Initialize the pre-processed dataset (PDS)

PDS= $\{PDS_1, PDS_2, \dots . PDS_n\}$ //n refers to number data from pre-processed dataset

Step 2:For F = 1 to number of each feature (N)

Compute weight (w_d) for each feature data

Calculate features subset size randomly

If Compute Possibility for each feature

Select feature with the highest possibility in a subset

Update each feature weights

End if

End for

Step 3: Return← D_w

Stop

The proposed Feature Selection using Relative Drug Intensive Weight (RDIW) consists of four steps: First, create a complete possibility with features updated. Relation of aspect weight. Then, disassemble the map using the community discovery algorithm. Clustering algorithm to find better partition. Finally, select a representative feature from each cluster and create a subset of the part. Let assume F refers to features.

3.3 Radial Subset Clustering Feature Selection (RSCFS)

In this phase is used to selects finest features of drug recommendation to improve classification prediction performance**.** A Radial Subset Clustering Feature Selection (RSCFS) method based on a three-step filter model. First, the inappropriate features are removed. Then, an automated feature clustering algorithm divides the feature set into multiple feature redundancy or related clusters. Finally, select a feature in the order of the subset section.

Algorithm steps

Input: Drug intensive weight

Output: Mutual features MIFS

Begin

Step 1: Initialize the Drug-intensive weight D_w dataset

Step 2: Process the intensive cooperative feature from both class Cif→(Lr∩ Br)

Step 3: Assimilate Radial basis decision → Intensification Range of values (Cif)

Mutual intensive rate (Mir→Nodes (Btree))

Generate cluster index to group constraint weight Cif(Mir)

For each decision for F features in I number of mean relevance feature

Rn→F (i) algorithm on each decision with marginal weight and features.

End for

Step 4: Compute the Br rate Lr rate mean depth Cif→Mir

Step 5: Extract the Correspondence feature weight Cfs

Return Cfs→ closest F(i) at the sum of mean depth relevance feature

End if

Stop

The above algorithm steps efficiently select the best features of drug compound molecules to reduce dataset dimension. Each group selects a function, data where used features were extracted from system image data sets. Compared to the seven feature selection algorithms, the results obtained show the efficiency and effectiveness of the system. The selective finest features are drug name, useful drug count, name of the condition. Let assume Cif refers to cooperative features, Mir presents mutual intensive rate, Rn refers to relevant features, Br refers to best features, and Lr refers to the least features.

3.4 Intra Segment Subset Feature Pattern Theory (ISSFPT)

Infra Segment Subset Feature Pattern Theory (ISSFPT) identifies important features and assigns necessary weight to each element from the feature selection dataset. In this phase, after feature selection dataset trained into ISSFPT technique for the proper weight of drug to improve the classification. Infra Segment Subset Feature Pattern Theory (ISSFPT) will lead to inoperable information. Provide solutions to descriptive ISSFPT issues and classify features as feature functions based on drugs features extraction theory.

Algorithm steps

Input: Mutual features MIFS

Output: Closest feature weights (Cfs)

Begin

Step 1: Initialize the Mutual features MIFS

Step 2: Compute for each feature weights$MIFS_i i$=1,2,….n

Step 3: Compute for each features drug important (I_d) class //I_d denote drug important features

While the important features satisfied do

For each$i$=1 do n then // i refers to iteration and n refers to number data

Calculate finest feature attributes fromCfs$_i$ in n do

End for

End while

Step 4: Calculate important feature weight I_w

For each$i$=1 do n then

$I_w = \sum_n^{i=1} \beta_i T_d(MIFS)c_l$ // β_i refers to feature optimal weight from training mutual features $T_d(MIFS)$ and c_l refers to class label of i ($c_l \in [-1, +1]$).

End for

Step 5: Update the closest feature weights Cfs

Step 6: Return←Cfs

Stop

The above algorithm step provides efficient weight features of drugs from the Mutual features MIFS dataset. Where,Cfs-Closest feature weights, an essential aspect of this method is that it is possible to accurately determine the discriminating clues of the feature subgroup with the most significant interaction and independence between features.

3.5 ReLU Activation Function

An improved Revised Linear Units (ReLU) section correction implementation function is proposed. On regulating deep spectral neural classification (DSNC) with ReLU in the hidden layer. It plays the role of pushing the input of ReLU to zero during the learning process by introducing sparsity in the input of ReLU.. Therefore, it is expected that the unwanted function decrease in ReLU output may be prevented. This is similar to the effect of volume feature weights normalization. It improves the generalization of the trained network.

$X_s = \begin{cases} Y_r \\ 0 \end{cases}$ if $Y_r \geq 0; Y_r < 0$ (1)

Where, $X,_s$ Y_r the relationship between weights and the proposed method and modification of modules or change of implementation functions. Detailed experiments check the effectiveness of this method. ReLU input may reduce the unwanted increase in ReLU output, and it has a similar effect to block normalization.

f_{Relu} (h_r, s) = max $(0, h_r, s)$ (2)

In addition,h_r, s ReLU's unwanted negative input variables can be minimized. This is expected to enhance the generalization ability of uniform weight loss.

3.6 Deep Spectral Neural Classification (DSNC)

In this phase, Radial Subset Clustering Feature Selection (RSCFS) dataset is trained into the DSNC algorithm to classify the recommendation of the drug. The marginalized features are effectively progressed to improve the classification fed into the classifier based on marginal weightage. The hidden layer performs the ReLuactivation function to predict the result and accuracy and reduce the computational complexity of the traditional algorithms.

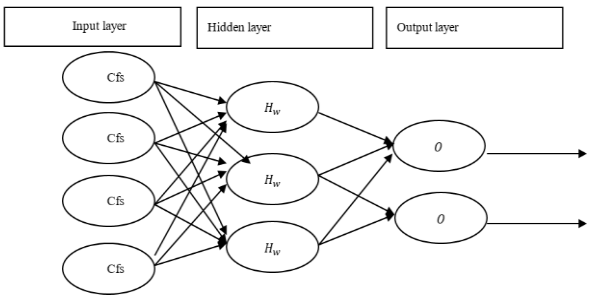

Figure 3: Process of Deep Spectral Neural Classification with hidden layer

Deep Spectral Neural Classification (DSNC) algorithm contains three layers, there are first input layers, second hidden layer, and third layer output layer. The first layer is the feature selection data fed into the input layer process supplied into hidden layer calculated the drug weights then fed into output layer to predict the drug recommendation using ReLu activation function present in figure 3.

Algorithm steps

Input: Closest feature weights (Cfs)

Output: Classification result as drug recommending (D_R)

Begin

Step 1: Initialize the closest weights (Cfs)

Step 2: Manage Each node n

Step 3: Compute the cluster feature (Fcl) as Frcids→ Fc11,Fcl2…

Feature process the ordered List Ord→{L1, L2….} to input layer

Step 4: Compute the decision classifier neural counter on the hidden layer

For initializing the value i=1 and j=1 asset count

For i=0 to n to fix marginal weightage class as threshold class

Compute the coefficient value of i→j as random point p

Spread cluster featured value (Scv→ i to j)

Initialized to set definite value P threshold margin

Set the controlling value of fitness value P to each trained node

End for

End For

Step 5: Evaluate the closest feature weights and ReLU to each logical node in i and j iterations

Step 6: While the termination criteria are not satisfied, gain (G) from fitness

G < Max Generation do

End while

Step 7: For Condition evaluate to feature weight terminate whether not Gain value <max generate value feature to the threshold value

If will G→max do

Check for i=1: n (compute as all n tasks to pread features)

Realizedfor j=1: n (n tasks) do

Check if (Xj<Xj), move task i towards j;

Return class

End for

Fin fi check weightage correlation along neurons

Compute the marginal weight of the class to split category by neighbor class

End if

Return Threshold margined class

Step 8: Output drug recommendation categorized class

Step 9: Return←D_R

Stop

The above algorithm steps gives efficient drug recommendation $D_R$prediction result using Deep Spectral Neural Classification (DSNC) with ReLu activation function. Let assume Xj refers to assetcount. Finally, predicted types projected to multi-data point closeness measure the weightage of data points, centralized to Neighbor clusters. This will improve the classification accuracy. The proposed algorithm effectively classifies the drug recommendation process.

4. Results and Discussion

This section describes the experimental setup process to evaluate the proposed Deep Spectral Neural Classification (DSNC) algorithm, and previous algorithms are Long Short-Term Memory Network (LSTM), Causal Inference-Probabilistic Matrix Factorization (CIPMF), and Graph Regularized Matrix Factorization (GRMF). The Drug Indications Database is collected from an online UCI repository. It contains 64 attributes: drug id, drug name, compound molecules (Chemical reaction) level, chemical id, and so on.

Table 1: Parameters settings

Parameters items	Values
Language	Python
Tool	Anaconda
Name of the dataset	Healthcare Drug Indications Database
Number of records	500000

The above table 1 shows the parameters settings for implantation of the proposed and existing methods carried out in Jupiter notebook in an anaconda environment. The total data is split into 70% training data and 30% test data. The confusion matrix is used to calculate all the parameters such as precision, recall, false rate, and prediction performance.

Table 2: Analysis of Drug Success Rate Prediction performance

Drug Success Rate Prediction Performance in %			
Methods	100000 Records	200000 Records	500000 Records
CIPMF	72	75	78
GRMF	76	80	83
LSTM	80	83	87
DSNC	85	87	93

Above table 2 presents the comparison result of success rate prediction in %. The proposed algorithm has high performance compared with the previous algorithm.

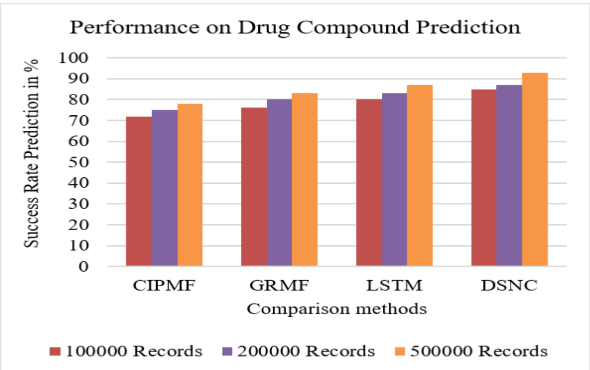

Figure 4: Analysis of Drug success rate prediction

Figure 4 presents an analysis of drug success rate prediction performance in percentage. The x-axis shows comparison methods, and the y-axis presents prediction performance. The proposed Deep Spectral Neural Classification (DSNC) algorithm obtained prediction result is 93% for 500000 Records. Similarly, the existing algorithms are Long Short-Term Memory Network (LSTM) result is 87%, and Causal Inference-Probabilistic Matrix Factorization (CIPMF) result is 78%, and Graph Regularized Matrix Factorization (GRMF) 83% for 500000 Records.

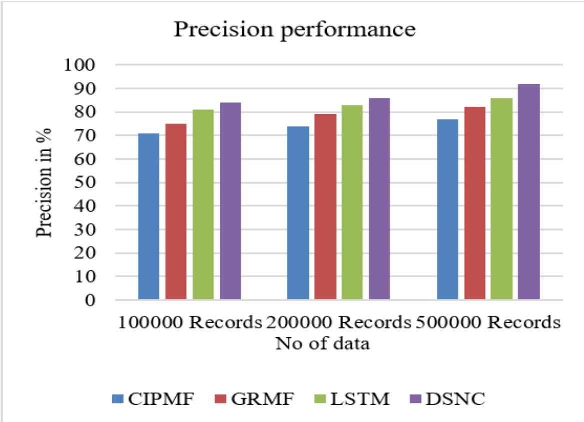

Figure 5: Analysis of Precision performance

Figure 5 illustrates the analysis of precision performance the recommended and existing results comparison. The x-axis shows the number of data, and the y-axis shows precision performance in percentage. The proposed DSNC algorithm precision result has 92% for 500000 Records, similarly, the existing algorithms are CIPMF precision has 77%, FRMF algorithm precision has 82%, and LSTM precision has 86% for 500000 Records.

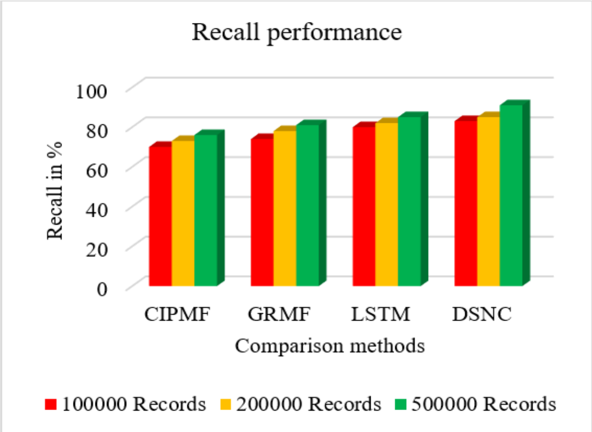

Figure 6: Analysis of Recall performance

Analysis of recall performance the proposed and existing algorithm comparison results are presented in figure 6. The proposed Deep Spectral Neural Classification (DSNC) algorithm recall result has 91%, likewise, the existing algorithm results are CIMPF algorithm has 76%, GRMF algorithm has 81%, and LSTM has 85%.

Table 3: Analysis of false rate performance

The false rate in %			
Methods	100000 Records	200000 Records	500000 Records
CIPMF	28	25	22
GRMF	24	20	17
LSTM	20	17	13
DSNC	15	13	7

Analysis of false rate performance comparison result present in table 3. The proposed algorithm provides low false classification performance.

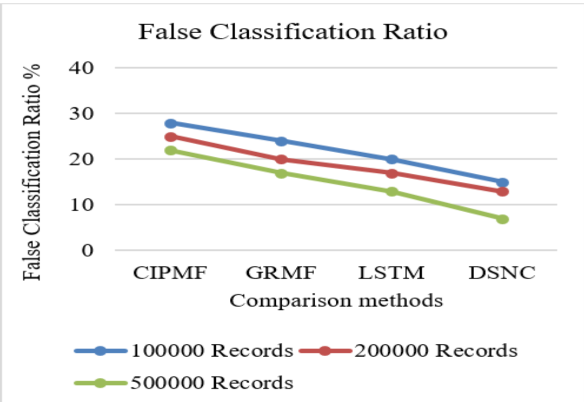

Figure 7: Analysis of false classification ratio performance

Analysis of false rate classification performance results is shown in figure 7. The proposed Deep Spectral Neural Classification (DSNC)false rate performance is 7%. Similarly, the existing algorithm is LSTM has 13%, GRMF has 17%, and CIPMF has 22%.

Table 4: Analysis of time complexity performance

Time Complexity performance in sec			
Methods	100000 Records	200000 Records	500000 Records
CIPMF	57	77	91
GRMF	44	65	79
LSTM	36	54	70
DSNC	20	25	31

Above, table 4 presents the time complexity for drug success rate prediction performance. The proposed algorithm provides low time compared with existing algorithms.

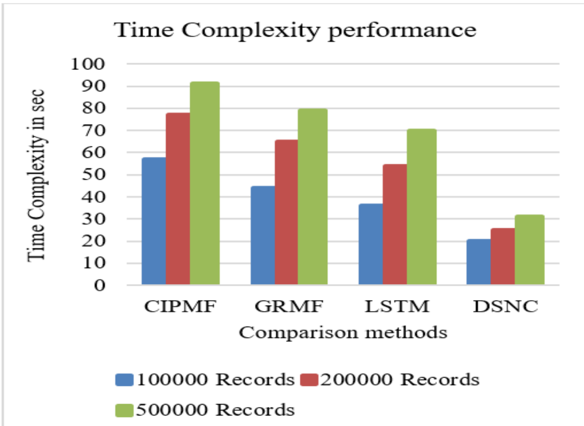

Figure 8: Analysis of Time Complexity performance

Analysis of time complexity performance results is shown in figure 8. The proposed Deep Spectral Neural Classification (DSNC) drug success rate prediction time complexity result is 31sec for 500000 Records. Similarly, the existing algorithms are Long Short-Term Memory Network (LSTM) drug success rate prediction time complexity is 70sec, Causal Inference-Probabilistic

Matrix Factorization (CIPMF) drug success rate prediction time complexity result is 91sec, and Graph Regularized Matrix Factorization (GRMF) 79sec for 500000 Records.

5. Conclusion

The proposed produce high performance by performing the analysis of drug compound analysis. This achieves the best intent values from the paramedical dataset by observing relational patterns. This relative drug compound analysis returns the mutual dependencies feature weights by finding the molecule substances related to drug compensates for identifying the features. Then the Radial subset clustering feature reduces the dimension of the features to create relative patterns. Then the classifier in Deep Spectral Neural Classification (DSNC) for predicting higher intensive rate based on mutual relation. This proposed system achieves high-performance through the confusion matrix weights probability in better performance than other methods. This system highly supports relational drug recommendation as well as other methods in drug compounds suggestion**.** The Proposed DSNC algorithm provides results are drug compound success rate prediction is 93%, recall is 91%, precision is 92%, false classification result is 7%, and time complexity result is 31sec. The proposed algorithm provides high accuracy performance with low time complexity and false rate performance.

References

A.Ezzat, P. Zhao, M. Wu et al., "Drug-target interaction prediction with graph regularized matrix factorization", *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 14, no. 3, pp. 646-656, 2017.

A.Wen, X. Sun, K. Yu, Y. Wu, J. Zhang and Z. Yuan, "Drug-Drug Interaction Extraction using Pre-training Model of Enhanced Entity Information," 2020 IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) pp. 527-532, doi: 10.1109/iThings-GreenCom-CPSCom-SmartData-Cybermatics50389.2020.00094.

D. Afdhal, K. W. Ananta and W. S. Hartono, "Adverse Drug Reactions Prediction Using Multi-label Linear Discriminant Analysis and Multi-label Learning," 2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Depok, Indonesia, 2020, pp. 69-76, doi: 10.1109/ICACSIS51025.2020.9263166.

F. Ali *et al.*, "Merged Ontology and SVM-Based Information Extraction and Recommendation System for Social Robots," in *IEEE Access*, vol. 5, pp. 12364-12379, 2017, doi: 10.1109/ACCESS.2017.2718038.

G. Jin and S. T. C. Wong, "Toward better drug repositioning: Prioritizing and integrating existing methods into efficient pipelines", *Drug Discovery Today*, vol. 19, no. 5, pp. 637-644, 2014.

H. Luo, J. Wang, C. Yan, M. Li, F. -X. Wu and Y. Pan, "A Novel Drug Repositioning Approach Based on Collaborative Metric Learning," in *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 2, pp. 463-471, 1 March-April 2021, doi: 10.1109/TCBB.2019.2926453.

H. Luo, M. Li, S. Wang et al., "Computational drug repositioning using low-rank matrix approximation and randomized algorithms", *Bioinf.*, vol. 34, no. 11, pp. 1904-1912, 2018.

H. Xue, J. Li, H. Xie et al., "Review of drug repositioning approaches and resources", Int. J. Biol. Sci., vol. 14, no. 10, pp. 1232-1244, 2018.

J. Yang, Z. Li, X. Fan et al., "Drug-disease association and drug-repositioning predictions in complex diseases using causal inference-probabilistic matrix factorization", *J. Chemical Inf. Model.*, vol. 54, no. 9, pp. 2562-2569, 2014.

J. Y. Ryu, H. U. Kim and S. Y. Lee, "Deep learning improves prediction of drug-drug and drug-food interactions", *Proc. Nat. Acad. Sci. United States America*, vol. 115, no. 18, pp. E4304-E4311, 2018.

Koras, K., Juraeva, D., Kreis, J. et al. Feature selection strategies for drug sensitivity prediction. Sci Rep 10, 9377 (2020).

https://doi.org/10.1038/s41598-020-65927-9

M. Fedorová, D. Perduková, Z. Pirnik, V. FedáK, O. Sukel and P. Sanjeevikumar, "The Fuzzy System as a Promising Tool for Drugs Selection in Medical Practice," in *IEEE Access*, vol. 6, pp. 27294-27301, 2018, doi: 10.1109/ACCESS.2018.2831282.

M. L. Shahreza, N. Ghadiri, S. R. Mousavi et al., "Heter-LP: A heterogeneous label propagation algorithm and its application in drug repositioning", *J. Biomed. Informat.*, vol. 68, pp. 167-183, 2017.

M. Tsubaki, K. Tomii, J. Sese, Compound-protein interaction prediction with end-to-end learning of neural networks for graphs and sequences, Bioinformatics 35 (2) (2019) 309–318, http://dx.doi.org/10.1093/bioinformatics/bty535.

N.J. Perualila-Tan, Z. Shkedy, W. Talloen, H.W. Göhlmann, M.V. Van Moerbeke, A. Kasim, Weighted similarity-based clustering of chemical structures and bioactivity data in early drug discovery, J. Bioinform. Comput. Biol. (2016) http://dx.doi.org/10.1142/S0219720016500189.

Shengyu Liu, Buzhou Tang, Qingcai Chen, Xiaolong Wang, Xiaoming Fan, "Feature Engineering for Drug Name Recognition in Biomedical Texts: Feature Conjunction and Feature Selection", Computational and Mathematical Methods in Medicine, vol. 2015.

SwathikClarancia Peter, Jaspreet Kaur Dhanjal, Vidhi Malik, NavaneethanRadhakrishnan, MannuJayakanthan, DuraiSundar. Quantitative Structure-Activity Relationship (QSAR): Modeling Approaches to Biological Applications. **2019**,,, 661-676. https://doi.org/10.1016/B978-0-12-809633-8.20197-0

T. Chen, P. Su, C. Shang, R. Hill, H. Zhang and Q. Shen, "Sentiment Classification of Drug Reviews Using Fuzzy-rough Feature Selection," 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), New Orleans, LA, USA, 2019, pp. 1-6, doi: 10.1109/FUZZ-IEEE.2019.8858916.

T. He, M. Heidemeyer, F. Ban, A. Cherkasov, M. Ester, SimBoost: a read-across approach for predicting drug-target binding affinities using gradient boosting machines, J. Cheminform. 9 (1) (2017) 1–14, http://dx.doi.org/10.1186/s13321- 017-0209-z.

T. Pahikkala, A. Airola, S. Pietilä, S. Shakyawar, A. Szwajda, J. Tang, T. Aittokallio, Toward more realistic drug-target interaction predictions, Brief. Bioinform. 16 (2) (2015) 325–337, http://dx.doi.org/10.1093/bib/bbu010.