PREDICTION OF DRUGS FOR DISEASES

B.Nithya[1], G.Anitha[2]*

ABSTRACT

Data mining techniques are applied in huge databases to find out the useful hidden patterns. There are different data mining approaches used for analysis, diagnosis, or prediction of diseases. After the prediction of disease, it is more important to recommend a drug based on the stage of the disease to a patient. There is no use of prediction of disease when a proper drug is recommended to the patient. Generally, in the medicinal field, there are two types of drugs are available such as generic name medicines and brand name medicines. But there is less number of researches has been carried out in the area of prediction of a drug for a disease. Various types of survey are present in drug recommendation techniques based on drug-disease interaction, drug-target interaction and side effects. It examines various techniques evaluated by previous researchers. The shortcomings of those methods are then discussed in order to identify advancements in drug recommendation methods.

Keywords: ANN, RNN, regression techniques, semi-supervised, drug, diseases, prediction

I. INTRODUCTION

From massive databases, useful hidden patterns have been found using data mining [1]. Different methods are frequently used to solve data mining difficulties such as machine learning, soft computing, data visualization, classifications and regression techniques. While some research projects are carried out in this area, they are all concentrated on a small number of techniques for the analysis, diagnosis, or prediction of diseases utilizing various tools and methodologies. But there is less number of researches has been carried out in the area of prediction of drug for a disease. In medicinal field, there are two types of drugs are available such as generic and brand name medicines. Drugs that are sold under a generic name are identical to their brand-name counterparts in terms of dose, intended use, mode of administration, hazards, safety, and strength. In other words, they have the same pharmacological effects as their brand-name equivalent.

counterparts. Patients are given drugs based on the disease's stage.

This research work is focused on the prediction of drug for treating Several diseases, including hypertension, diabetes, and chronic heart failure using some data mining techniques. Initially, multiple attributes of drugs and patients are extracted. Along with these features, hidden relationship among features and statistical features are extracted by using Hidden Markov Model (HMM). Some of the extracted features are temporal features, those feature values are varying with respect to time. The optimal time period for temporal features is selected by using krill herd algorithm. Then, these features are used in Artificial Neural Network (ANN) which predicts drugs for diseases. The side effects due to consuming generic and branded medicines are analyzed using auto encoders which is a deep learning technique. Moreover, drug reaction has the same deep learning technique, medications' biological, chemical, and phenotypic properties are also examined using auto encoders. Finally, prediction of drug reaction is improved by using advanced deep learning techniques such as Recurrent Neural Network (RNN) and Hierarchical Fused Fuzzy Deep Neural Network.

[1]Department of Computer Science

[2]Department of Computer Application

Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India

*Corresponding Author

II. LITERATURE SURVEY

By employing electronic clinical data that indicates the end-point physiological consequences of pharmacological effects on human biological activities, a novel computational framework [2] was proposed to draw conclusions for alternate indications of commercially available pharmaceuticals. The idea of complementarities between clinical illness markers and clinical medication effects was applied in this study. Using this paradigm, two methodologies—statistical analysis and literature mining were applied to create disease-related clinical variable vectors and drug-related variable vectors. Each disease-drug pair was then given a repositioning potential score through the calculation of complementarities and relationship between clinical states of disease signatures and clinical effects. of drugs. However, small sample size was used in this framework.

Based on claims information from[3] Blue Cross Blue Shield of Texas, individuals who received at least one prescription for a diabetes medication between 2008 and 2011 were split into a training group (90% of patients) and a test set (10%). It was explained how to use cSPADE to track the temporal trends of diabetic medication prescriptions. These mined patterns, which were represented as digraphs, were used to deduce the temporal correlations. Finally, rules that predict the following diabetic medicine prescribed for a test group of patients are generated using the knowledge base of patterns that have been mined. However, this takes a lot of time and an expensive process.

Using information fusion and aggregation, a computational technique known as Scored Mean Kernel Fusion (SMKF) [4] was introduced for the prediction of drug-disease relationships. There are numerous features and a variety of data sources that can help forecast unknown drug-disease interactions and produce more accurate and trustworthy findings. By combining data from different sources and employing aggregation operators, this information can be mined collectively. Therefore, in the computational method, high-level features were created using the feature fusion method. The average aggregation operator known as score mean was scored utilising the SMKF approach. This strategy systematically incorporated many characteristics of medications or diseases at two levels—drug-drug level and drug-disease level—to anticipate novel drug indications. Drug illness can be improved by utilising additional drug properties, such as the route feature.

For the purpose of predicting drug-disease interactions, a semi-supervised graph cut method and three-layer data integration [5] were created. Three layers of the heterogeneous data were combined hierarchically. A unique weighted drug-disease pair network with nodes representing drug-disease pairs with known or unknown therapeutic relationships and edges representing node-node relations that were weighted with the similarity score between two pairs was created. The best network graph cut was discovered because similar drug-disease combinations should have comparable treatment trajectories. It was thought that the drug-disease pair with an unknown relationship had a comparable therapeutic relationship to that of the other drugs in the same cut. A semi-supervised graph cut algorithm (SSGC) was created as a result to choose the best graph cut and determine the possibility

III. PROBLEM DEFINITION

The problems considered for the research work are follows:

- Prediction of proper drug with no or less side effects is required to care a patient effectively.
- Drug and disease relation is changing over time. So, optimal time is required for analyzing the temporal features for prediction of drug for disease.
- Drug reaction has high impact for prediction of drug which is required to analysis to improve the accuracy of drug prediction.
- Deep learning is a fully deterministic model which doesn't consider the data uncertainty reduction.

IV. .PROPOSEDCONTRIBUTIONS

4.1 Prediction of drugs for diseases with the help of side effect and patient physical attributes

In the first phase of the study, information about patients and drugs is gathered. This information includes the name of the drug, its combination of ingredients, size, shape, tablet coating, weight, surface area, disintegration time, tendency to swell, side effects, mode of administration (tablet, injection, or capsule), cost of the drug, class, dose, brand name, drug reaction, patient age, gender, weight, and height. In the collected features there will be a hidden relation among features. Such hidden relations are extracted by using Hidden Markov Model [6]. Moreover, the statistical features such as mean, standard deviation, skew and kurtosis are extracted using the HMM. HMM is a sequence model which is described as the calculation of the probabilities that a feature is extracted by a model or sequence of model. The extracted features are given as input to the Deep Neural Network (DNN). There are multiple diseases that can be treated by this method, including diabetics, hypertension, and chronic kidney disease.

4.2 Feature selection and extraction for Prediction of drugs for diseases

In the second work of the research, feature selection and extraction method for prediction of drugs for diseases is proposed. Some of the collected features are temporal features. The temporal features are the one which varies their values with respect to time. To increase prediction accuracy, temporal characteristic time periods are adjusted for multiple diseases. The time period is optimized by using Krill Herd (KH) optimization algorithm [7]. The KH algorithm is based on a simulation of krill individuals flocking together. The aim function for krill migration is thought to be the lowest lengths that each individual krill must travel to reach food and the herd's highest density. Three key factors—foraging activity, movement caused by the presence of other individuals, and random diffusion—formulate the krill individuals' time-dependent positions. At the optimal time period, the value of temporal features is considered for drug prediction which is given as input to DNN along with the extracted other features. Finally, the DNN is used to predict the drug for multiple diseases.

4.3 Differentiate branded and generic drug based on reaction, side effect

In the third work of the research, a deep learning method called as Auto encoders (AE) [8] is introduced to analysis various features of drugs. This method examines a number of aspects, including the biological, chemical, and phenotypic characteristics of medications as well as adverse effects brought on by drug use. Targets, transporters, and enzymes are examples of biological features of pharmaceuticals. Substructure fingerprints are an example of chemical properties. Phenotypic properties include side effects and therapeutic indications.AE, a feed-forward network that learns to reconstruct the input information, analyses these features. A set of recognition weights are used to train AE to encode the input into a feature space. After that, a set of generative weights is used to approximate recreate the features (codes). The majority of the time, the generative weights are acquired from the encoder's unrolled weights initially, followed by a finely tailored phase. The encoder reduces the size of x to the specified number of codes (i.e., mapping to a hidden representation). The input through the decoder is then rebuilt using the codes. The algorithm is unsupervised because labelled training data is not necessary. Thus the Auto Encoderanalyze the various features of drugs and predict the drug for multiple diseases effectively.

4.4 Improving reaction prediction using advanced deep learning

Recurrent Neural Networks (RNN) and Hierarchical Fused Fuzzy Deep Neural Networks are two examples of sophisticated deep learning employed in the fourth research study to improve drug reaction prediction. Three hidden layers are sandwiched between an input layer and an output layer in the feed-forward multi-layer perceptron that makes

up the RNN. The RNN's purpose is to replicate the input data pattern at the output layer with the least amount of error possible through training. Information is extracted from both fuzzy and neural representations by the hierarchically fused fuzzy deep neural network. The information gained from these two viewpoints is combined in a fusion layer to create the final drug prediction representation. In details, fuzzy representation reduces the uncertainties and neural representation removes the noises in the original data.

V. CONCLUSION

In this paper, a detailed survey on drug recommendation based on data mining techniques is presented. It is obvious all researchers have tried in different techniques to recommend the proper drug for a disease to improve the health of a patient. The discussed drug recommendation techniques provide the recent development techniques for drug recommendation based on drug-disease interactions, drug-target interactions, side effects are analyzed by analyzing novel ideas incorporated in them. The analysis of these techniques provides the better understanding of these steps involved in each process thus increasing the scope for finding the efficient techniques to achieve better performance. Based on the analysis, Network Topological Similarity-based Inference Method, Network Topological Similarity-based Classification Method has better performance than the other techniques. This survey helps in deriving the motivation for our future researches as well.

REFERENCES

1. Bhavya, & Mittal, P. “Comparative Analysis of Data Reduction model for Diabetes”, International Journal of Computer Science and Mobile Computing, 4(7), 315-324, 2015.

2. Jang, D., Lee, S., Lee, J., Kim, K., & Lee, D. (2016). Inferring new drug indications using the complementarity between clinical disease signatures and drug effects. Journal of biomedical informatics, 59, 248-257.

3. Wright, A. P., Wright, A. T., McCoy, A. B., & Sittig, D. F. (2015). The use of sequential pattern mining to predict next prescribed medications. Journal of biomedical informatics, 53, 73-80.

4. Moghadam, H., Rahgozar, M., & Gharaghani, S. (2016). Scoring multiple features to predict drug disease associations using information fusion and aggregation. SAR and QSAR in Environmental Research, 27(8), 609-628.

5. Wu, G., Liu, J., & Wang, C. (2017). Predicting drug-disease interactions by semi-supervised graph cut algorithm and three-layer data integration. BMC medical genomics, 10(5), 79.

6. Salcedo, F. J., Díaz-Verdejo, J., & Segura, J. C. (2007). Features Extraction for Music Notes Recognition Using Hidden Markov Models. In Proceedings of the International Conference on Signal Processing and Multimedia Applications SIGMAP, pag184-191.

7. Gandomi, A. H., & Alavi, A. H. (2012). Krill herd: a new bio-inspired optimization algorithm. Communications in Nonlinear Science and Numerical Simulation, 17(12), 4831-4845.

8. Yousefi-Azar, M., Varadharajan, V., Hamey, L., & Tupakula, U. (2017, May). Autoencoder-based feature learning for cyber security applications. In Neural Networks (IJCNN), 2017 International Joint Conference on (pp. 3854-3861). IEEE.

9. Osipov, V., & Osipova, M. (2018). Space-time signal binding in recurrent neural networks with controlled elements. Neurocomputing.

10. Deng, Y., Ren, Z., Kong, Y., Bao, F., & Dai, Q. (2017). A hierarchical fused fuzzy deep neural network for data classification. IEEE Transactions on Fuzzy Systems, 25(4), 1006-1012.