**International Journal of Research Publication and Reviews**

Journal homepage: www.ijrpr.com ISSN 2582-7421

# Deep Learning Model for Drug Recommendation System

*[a] S. K. Avanthi, [b] P. Vinod Kumar, [c] P. Manikanta, [d] P. L. Tirumalesh, [e] V. L. Siva Prasad*

*[a,b,c,,d e] B.Tech., Student, ECE Dept, GMRIT(A), Rajam, A.P, India*

**ABSTRACT**

Online recommendation systems for drugs, medical professionals, and hospitals are gaining attention among the public to acquire health care services online. Online recommendation systems are becoming more crucial for drugs, hospitals, doctors, and pharmaceuticals. Due to a lack of availability, people began taking medication on their own without proper consultation, which made their health situation worse than usual. Deep learning has become increasingly useful in a variety of recent applications. A drug recommendation system using deep learning models that is capable of accurately predicting medicine based on the user's input of diseases or symptoms. Especially in times of medical emergency, information like this could be helpful in safe medications to patients. Artificial Neural Networks (ANN), and Convolution Neural Networks(CNN) are some deep learning algorithms that will be used to recommend medicine. The results of the models will be compared with accuracy classification tools, which are precision, recall, and f1 score.

Keywords: *Deep learning, Convolutional neural network, Artificial Neural Network, Classification.*

## 1. INTRODUCTION

Experts regularly make mistakes when prescribing medication—more than 35% of the time—because they have only access to a restricted quantity of information. Choosing the correct medicine is crucial for everyone who requires healthcare providers with a vast understanding of microscopic organisms, antibacterial medications, and patients. The best recommender framework is a common system that makes an item recommendation to the user based on their Diseases.

Generally, people with age above 60 consume two to three drugs at a time so which leads to unwanted effects called drug-drug interaction. In this case, a drug-drug interaction takes place, resulting in several adverse drug reactions (ADRs) that could result in harm or even death. Therefore, in order to prevent unexpected adverse drug reactions (ADRs) and enhance synergistic advantages while treating a condition to some extent, it is crucial to efficiently detect possible Drug-Drug Interactions (DDIs).To get chemical, and pharmacological data on drugs the Drug Bank database is used.

In deep learning, we don't need to explicitly program everything. In the olden days, we did not have that much processing and pre-processing power and a lot of data. In the last 15 years, processing power increased rapidly, and the technologies of deep learning and machine learning came into the frame. Deep learning can be considered a subset of machine learning and it is based on learning and improving on its own by examining computer algorithms. Deep learning has been deployed in an enormous number of applications. The Deep Learning models were successfully used to solve many critical and large sets of data problems based on Artificial Intelligence. Most study fields that produce some kind of data, whether it be text, images, audio, or video, can benefit from using deep learning techniques. It is relevant to drug discovery, genetics, natural language processing (NLP), and computer vision. The biggest benefit of deep learning architectures is their ability to interact with models and nonlinear relationships in data to produce better representations than the learning process itself can. Convolutional neural networks (CNNs) have been used to extract local residue patterns and forecast binding scores from drug and protein sequence data. However, rather than the elements themselves, a protein's or drug's properties depend on the sequence in which they are arranged.

Deep learning is one of the most important AI methods, particularly when dealing with images or movies. In this talk, I will introduce deep learning, with a particular focus on people who are experts in a biomedical area but not in AI. I will discuss how it works, at a nontechnical but functional level, the differences between traditional computational approaches and deep learning, how networks are trained, and which areas they are ineffective at, excel at, and will hopefully excel at in the future. The goal is to build an artificial structure known as an artificial neural network with computer nodes that function similarly to these brain neurons. In the hidden layer, there may be a large number of linked neurons between the neurons used for input and output values. Numerous applications have found use for machine learning, and creative work for automation is on the rise. This essay aims to offer a heap-reducing drug recommendation method. In this study, we developed a drug recommendation system that uses patient reviews to predict drug-drug interactions using a variety of factorization techniques like deep learning and ANN (artificial neural networks). The system can also suggest the best drug for a given disease by utilising a variety of classification algorithms. Precision, recall, accuracy, and AUC score were used to assess the anticipated medications.

## 2. Related Work

[1]. The process of selecting medications for the disease is powered by increasing amounts of information from already-existing chemical libraries and data banks. To research a medicine's chemical content, drug banks are useful. Neural network-based models that is Graph Neural Network (GNN) is used in the paper. The knowledge graphs is used to extract relevant information from data. GNN is employed because it can find node-to-node correlation. The initial step in medication design is target identification. Target-drug interactions can change how cells operate, leading to the development of effective treatments for disease. In order to fully utilize the algorithm and molecular graphs, the created model approaches are based on multimodal learning, which is a promising direction for better prediction in the drug development industry.

[2]. Health-related recommender systems can make an early diagnosis, anticipate disease, and offer the appropriate recommendations based on the patients' health status. The ANN model of artificial neural networks is utilized to treat infectious diseases. The dataset for COVID-19 treatment used in this study was compiled from a variety of sources, including drug banks, news articles, and published literature. To forecast drug adverse effects linked to drug names and clinical factors, drug attributes are retrieved. The error rate was compared when the number of hidden layers in the medicine recommendation system with stacked ANN was raised. Gender, age, weight, COVID-19, exercise habits, test results, country, and eating habits are among the features in the dataset. Results from the suggested system were 97.5% accurate.

[3]. Patients who are afflicted with several illnesses, in particular, are given safe pharmaceutical recommendations using the Safe Medical Recommendation (SMR) method. For drug-related data from Drug Bank, knowledge graphs are built. Real EMRs datasets-MIMIC-III is the dataset that was utilized in the model's implementation. Five or more drugs are routinely taken at once by patients with two or more disorders, posing major health hazards. Electronic medical records (EMRs) are utilized to help clinicians as a means of overcoming this and enabling them to make better clinical judgments. Using embedding, Safe Medicine Recommendation breaks down drug recommendations into a link prediction process that takes the patient's diagnosis and adverse drug reactions into account. Accuracy can be improved by incorporating more patient data, such as clinical outcomes.

[4]. Expert advice assists specialists in their research into the numerous chemical features of drugs during the COVID-19 epidemic, which aids in the discovery of the drug. Deep learning technology was used to build the graph neural network (GNN)+long short-term memory (LSTM)+generative adversarial network (GAN) expert recommendation. Each node must be represented by a low-dimensional state vector in the model's implementation in order to encode the structural information of the graph. The primary purpose of the output layer is to create a list of suggestions based on the anticipated outcomes. The GNN analyses the graph's attributes. The data from the GNN is fed into the LSTM network to follow the dynamic evolution of the graph. Then, we employ GAN to produce a superb recommendation. This model's accuracy is 89.89%.

[5]. Technologies for providing health recommendations have been developed to assist both patients and physicians in enhancing care, reducing errors, and saving time. 25 280 patient records from a real-world data set from the ruijin hospital were used. We coordinate Deep Neural Network (DNN) with existing medical information for the Diabetic medication recommendation system. The combination of Recurrent Neural Networks (RNNs) with graph knowledge can improve accuracy. This study suggests a DIMERS recommendation model that integrates clinicians' existing medical knowledge with bidirectional long short-term memory. Precision - 0.877 is the performance metric for the diabetic medication recommendation system.

## 3. Methodology:

A recommender framework is a common system that makes an item recommendation to the user based on their Diseases. Recommendation system have been developed to help patients and doctors improve services, reduce errors, and save time. By taking the patient symptoms the recommendation system recommends the drug that help the patient. To increase the safety of treating infectious diseases, drug recommendation systems using Convolutional Neural Network and stacked artificial neural network (ANN) models are used. The neural structure, which is organized into layers and is made up of hundreds of single artificial neurons connected with weights, is the basis of an ANN.We had implemented the convolutional neural network and combinations of layers in different models Embedding, Long Short-Term Memory (LSTM), Gated Recurrent Unit Network (GRU) and a Convolutional 1D layer (Conv1D).

### *Dataset:*

The dataset UCI machine learning Drug Review dataset is taken form kaggle. Previously, this dataset was used for the winter 2018 Kaggle University Club Hackathon and is now publicly available. There are two datasets train dataset and test dataset. The datasets train and test are textual datasets consists of seven columns: unique ID, drug Name, condition, review, rating, date, useful Count. The train dataset contains 161297 records and having 112329 unique patient reviews. The test dataset contains 53766 records and having 48280 unique patient reviews.

### *3.1 Dataset Pre-processing:*

This project is based on textual data. So, the pre-processing is done on textual data using natural language processing techniques. Text pre-processing is a method to clean the text data and make it ready to feed data to the model. Text data contains noise in various forms like emotions, punctuation, text in a different case and the data also contains the null values. When we talk about Human Language then, there are different ways to say the same thing, And this is only the main problem we have to deal with because machines will not understand words, they need numbers so we need to convert text to numbers in an efficient manner.

### *Removing NULL values:*

The real-world data often has a lot of missing values. The cause of missing values can be data corruption or failure to record data. The handling of missing data is very important during the pre-processing of the dataset. There many ways to handle missing data.

Here are some ways to handle missing data:

1. Deleting Rows with missing values.
2. Replacing with Mean/Median/Mode.
3. Assigning a Unique Category.
4. Predictingthe missing values.
5. Using Algorithms Which Support Missing Values.

In this project we had handled the missing data by deleting rows with missing values.

### *Tokenization:*

Tokenization is the process of dividing text into a list of tokens from a string of text. Tokenization involves cutting the raw text into manageable pieces. Tokenization divides the original text into tokens, which are words and sentences. These tokens support the construction of a model for natural language processing or the understanding of the context. By examining the word order in the text, tokenization aids in comprehending the text's meaning.

### *Embedding:*

To implement word embeddings, the Keras library contains a layer called Embedding(). The embedding layer is implemented in the form of a class in Keras and is normally used as a first layer in the sequential model for NLP tasks. The embedding layer can be used to perform three tasks. It can be used to learn word embeddings and save the resulting model. It can be used to learn the word embeddings in addition to performing the Natural Language Processing (NLP) tasks such as text classification, sentiment analysis, etc. It can be used to load pre trained word embeddings and use them in a new model. There are three parameters to be passed to the embedding layer. The first parameter in the embedding layer is the size of the vocabulary or the total number of unique words in a corpus. The second parameter is the number of the dimensions for each word vector. For instance, if you want each word vector to have 32 dimensions, you will specify 32 as the second parameter. And finally, the third parameter is the length of the input sentence.

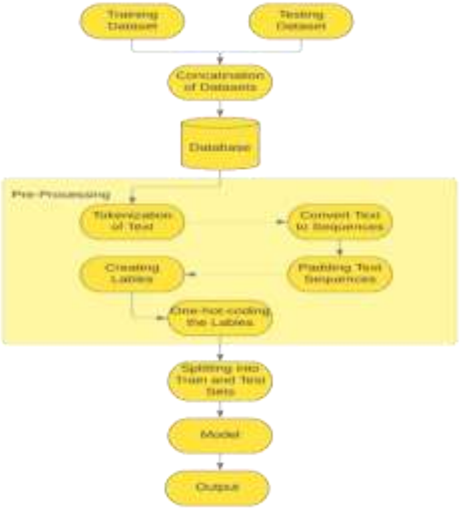
## *3.2 Our Model:*

### *Data Processing:*

The training and testing datasets are imported by using pandas package to process the raw data and making it suitable to the deep learning model. The both training and testing datasets are concatenated into one dataset. Then the records are shuffled. The dataset can be checked if there are any NULL values. The records with NULL values are removed. The dates format is converted into date time format. The Hexadecimal values '&#039' are in the review column replaced with the proper character single quote ('). Here the character single quote must be replaced so that natural language toolkit works efficiently.

The conditions are removed which are have <span> in it. Some stop words in the reviews are removed to clean the reviews. The labels for the drugs are created based on rating for the drug. If the rating of the drug is less than 5 then it is labelled as -1 because the drug will work with less efficiency. If the rating of the drug is greater or equal to 5 then it is labelled as 1 because the drug will work with high efficiency. The cleaned reviews are tokenized and converted into sequences. Tokenization is used to splitting up of larger review of text into smaller lines, words or even creating words for a non-English language. The padding is applied to the text sequences. Pad sequences is used to ensure that all sequences in a list have the same length by default this is done by padding 0 in the beginning of each sequence until each sequence has the same length as the longest sequence. The labels are converted into one hot categories lables by using to_categorical function. The function to_categorical converts a class vector (integers) to binary class matrix. The dataset is split into training and testing sets of sizes 80% and 20%.

Then the training set is given to the model for training. The flowchart of our proposed work is shown below-



### *3.3 Result:*

We have used three models CNN model, Artificial Neural Network(ANN)and Long short term memory(LSTM). Compared to ANN and CNN and LSTM models, CNN showed better performance with a testing accuracy 92% and ANN with 86% accuracy . The Accuracies of the models are compared in the comparison table which is shown in below Table

| **S.No** | **Model** | **Training Accuracy** | **Validation Accuracy** | **Testing Accuracy** |
|---|---|---|---|---|
| 1 | CNN | 98.40 | 90.38 | 92% |
| 2 | ANN | 80.07 | 82.3 | 86% |
| 3 | LSTM | 96.83 | 48.8 | 60.9% |

## 4. CONCLUSION & FUTURE SCOPE:

The dataset used by the UCI machine learning Drug Review is obtained from Kaggle. Word embedding is performed on the datset. It is an alternative method of preprocessing the data. This embedding can map words with comparable semantic meanings. employing two separate models, such as the Custom CNN model and ANN model, to then categorise. The accuracy of our custom CNN model was 92%, whereas that of the other models (ANN) was 86%. By taking into account more patient data in the future, this work's accuracy can be increased. It might also help with future studies to find new DDIs and their pharmacological results. By implementing this model and leveraging tools like Flask and Django for greater user interaction, the UX can also be enhanced.

### References

[1] Zeng, X., Tu, X., Liu, Y., Fu, X., &Su, Y. (2022). Toward better drug discovery with knowledge graph. Current opinion in structural biology, 72, 114-126.

[2] Bhimavarapu, U., Chintalapudi, N., &Battineni, G. (2022). A Fair and Safe Usage Drug Recommendation System in Medical Emergencies by a Stacked ANN. Algorithms, 15(6), 186.

[3] Gong, F., Wang, M., Wang, H., Wang, S., & Liu, M. (2021). SMR: medical knowledge graph embedding for safe medicine recommendation. Big Data Research, 23, 100174.

[4] Wang, H., & Le, Z. (2021). Expert recommendations based on link prediction during the COVID-19 outbreak. Scientometrics, 126(6), 4639-4658.

[5] Wedagu, M. A., Chen, D., Hussain, M. A. I., Gebremeskel, T., Orlando, M. T., & Manzoor, A. (2020, December). Medicine Recommendation System For Diabetes Using Prior Medical Knowledge. In Proceedings of the 2020 4th International Conference on Vision, Image and Signal Processing (pp. 1-5).

[6] Xiong, G., Yang, Z., Yi, J., Wang, N., Wang, L., Zhu, H., ... & Cao, D. (2022). DDInter: an online drug–drug interaction database towards improving clinical decision-making and patient safety. Nucleic acids research, 50(D1), D1200-D1207.

[7] Granda Morales LF, Valdiviezo-Diaz P, Reátegui R, Barba-Guaman L. Drug Recommendation System for Diabetes Using a Collaborative Filtering and Clustering Approach: Development and Performance Evaluation. J Med Internet Res. 2022 Jul 15;24(7):e37233. doi: 10.2196/37233. PMID: 35838763; PMCID: PMC9338420.

[8] L. Li and M. Cai, "Drug Target Prediction by Multi-View Low Rank Embedding," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 16, no. 5, pp. 1712-1721, 1 Sept.-Oct. 2020, doi: 10.1109/TCBB.2017.2706267.

[9] S. Garg, "Drug Recommendation System based on Sentiment Analysis of Drug Reviews using Machine Learning," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 175-181, doi: 10.1109/Confluence51648.2021.9377188.

[10] Liu, Shichao, et al. "Enhancing drug-drug interaction prediction using deep attention neural networks." IEEE/ACM Transactions on Computational Biology and Bioinformatics (2022).

[11] Shao, K., Zhang, Z., He, S., & Bo, X. (2020, November). DTIGCCN: Prediction of drug-target interactions based on GCN and CNN. In 2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI) (pp. 337-342). IEEE.

[12] Wang, Y., Chen, W., Pi, D., Yue, L., Wang, S., & Xu, M. (2021, August). Self-Supervised Adversarial Distribution Regularization for Medication Recommendation. In IJCAI (pp. 3134-3140).

[13] Rezaei, M. A., Li, Y., Wu, D., Li, X., & Li, C. (2020). Deep learning in drug design: protein-ligand binding affinity prediction. IEEE/ACM Transactions on Computational Biology and Bioinformatics.

[14] Monteiro, N. R., Ribeiro, B., &Arrais, J. (2020). Drug-target interaction prediction: end-to-end deep learning approach. IEEE/ACM transactions on computational biology and bioinformatics.

[15] Tan, Y., Kong, C., Yu, L., Li, P., Chen, C., Zheng, X., ... & Yang, C. (2022, August). 4SDrug: Symptom-based Set-to-set Small and Safe Drug Recommendation. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 3970-3980).