International Journal of Engineering Trends and Technology Volume 72 Issue 10, 140-148, October 2024
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V72I10P114

Original Article

Leveraging Text Mining for Drug Recommendation System

Sarojini Balakrishnan[1], D. Sobya[2]

[1]*Department of Computer Science, Avinashilingam Institute of Home Science and Higher Education for Women, Coimbatore, Tamil Nadu, India.*
[2]*Department of Computer Science & Engineering, Institute of Engineering and Management, University of Engineering and Management, Kolkata, West Bengal, India.*

[1]*Corresponding Author : sarojini_cs@avinuty.ac.in*

Received: 19 June 2024 Revised: 02 October 2024 Accepted: 14 October 2024 Published: 25 Ocotber 2024

Abstract *- The advent of social media has significantly empowered patients to share their medication experiences across various online platforms. These reviews reflect diverse sentiments, highlighting the positive and negative effects of the prescribed drugs on their health. Analyzing these user-generated reviews on social media can uncover latent details regarding the efficacy of the drugs, the possible side effects, and the patient's satisfaction level. These reviews also help other stakeholders, like pharmaceutical companies and healthcare professionals, gain valuable insights about the drug. Text mining techniques can be leveraged to examine these reviews and identify their associated sentiments. In this research, we develop a Drug Recommendation System using Machine Learning and Deep Learning models. The sentiments of the patients are analyzed to decide on the most suitable drug for a particular medical condition. The performances of these models were evaluated and compared using metrics - accuracy, precision, recall, and F1 score. Empirical results demonstrate that Bi-directional RNN and Light Gradient Boosting models outperformed other models taken for this study.*

Keywords *- Deep learning models, Drug review, Machine learning models, Text mining, Sentiment analysis.*

1. Introduction

Drug safety is a vital aspect of healthcare. Primarily, medical professionals must ensure that the drugs they prescribe are safe and have minimal side effects [1]. Drug safety is typically assessed using clinical trials and specific medical techniques. Alternatively, the patients, the key stakeholders, can provide valuable insights about the impact of their medications based on their personal experiences. Patients use social media and online healthcare forums such as RateMDs, WebMD, Ask a Patient, DrugLib.com, Drugs.com, MedHelp, and Daily Strength to express their emotions and opinions about their health conditions, the doctors, and the medications they consume. Patient's online reviews regarding the medicines they use offer a piece of first-hand information, encompassing not only the efficacy of the drugs but also the side effects. The reviews expressed by the patients may be of positive or negative tones based on the effects of a particular drug on their health conditions. Analyzing these reviews expressed by the patients is crucial for healthcare professionals and drug developers, as it provides essential insights about the drugs, which can significantly influence healthcare practices and drug development [2]. Sentiment analysis of drug reviews from various perspectives could demonstrate the efficacy of the drugs and unveil the potential risks [3,4]. It can also provide customized treatment plans that address patients' unique needs and preferences [5]. A drug recommender system can be developed based on patient reviews to suggest a suitable medicine for a specific medical condition to realize the above-said application. Text mining techniques are used to analyze these user-generated data - reviews, comments, and social media posts. They play a crucial role in uncovering hidden patterns, trends, and insights from large volumes of unstructured user-generated text data. The complex and varied expressions in the reviews posted by drug users can be analyzed using Natural Language Processing (NLP) techniques to identify the sentiments associated with these sentences [6,7]. This research aims to uncover critical insights into patient experiences and perceptions of medications embedded in drug reviews, potentially leading to breakthroughs in drug development and improved health outcomes. However, as with other sentiment analysis applications, it is essential to consider accuracy and bias when analyzing drug reviews. In this work, we develop a Drug Recommendation System that suggests drugs to patients based on their medical conditions. The Machine Learning (ML) and Deep Learning (DL) models taken up for this study are trained on the Drug Review dataset from the

UCI ML repository. We use NLP techniques to process and categorize the reviews in the dataset into positive and negative sentiments. To ensure that a comprehensive system is built, we compare the performance of multiple ML and DL models in the task of classifying the reviews. The ML models taken up for this comparative study are - Light Gradient-Boosting Machine (LGBM), Multinomial Naïve Bayes (MNB), Logistic Regression (LR), and Decision Tree (DT). The DL models taken up for this study are - Bidirectional Recurrent Neural Network (BRNN), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Deep Neural Networks (DNN). The performances of the classification models were evaluated using the metrics of accuracy, precision, recall, and F1 score. This article is structured in the following way: Section 2 discusses the existing works in the literature. Section 3 explains the methodology for building the proposed drug recommendation system, and Section 4 analyses its performance. The article is then concluded in Section 5.

2. Related Works

Many research studies in the literature focus on sentiment analysis of user-generated text in healthcare forums [8,9,10]. Sentiment analysis involves identifying the emotions conveyed in a text, focusing on understanding the overall tone of a statement rather than simply determining whether specific words carry positive or negative meanings [11]. By analyzing the emotional tone or attitude conveyed in these reviews, sentiment analysis can offer a comprehensive and varied viewpoint on the efficacy, side effects, and overall satisfaction regarding a specific product [12]. Sentiment analysis for drug reviews initially relied on rule-based methods [13] and used sentiment lexicons like SentiWordNet [14] to calculate a reviewer's overall polarity, which can be either positive or negative. However, drug reviews can often include personal opinions and subjective language that might not align with predefined rules or lexicons, resulting in inaccurate analysis and interpretations of user-generated content.

Also, misspelt words may not be adequately handled, which reduces the accuracy and efficiency of these systems [15,16]. Bobicev et al. [17] introduced the Bag of Words (BoW) approach that disclosed sentiments expressed in Twitter messages. Vijayaraghavan et al. [18] conducted a comprehensive study on the evaluation of drug reviews through text analysis and the rating scheme. Their main objective was to propose supervised ML models that could accurately classify pharmaceutical reviews into positive, neutral, or negative categories based on the text reviews. They employed Term Frequency-Inverse Document Frequency (TF-IDF) and Count Vectors (CV) and developed ML models that were specifically tailored to three common factors: depression, birth control, and pain. Zmandar et al. [19] demonstrated the use of NLP for financial document processing and emphasized the importance of domain-specific training. The proposed study uses multiple languages, generating word embeddings for the shortlisted ones. English embeddings improved performance on a sentiment analysis classification task on the Financial Phrase dataset, outperforming a standard GloVebased model. This research work provided a valuable reference for NLP implementation with multiple languages. Tianhua Chen et al. [20] proposed a technique for sentiment analysis of medication reviews using the Fuzzy-Rough Feature Selection (FRFS) method. The approach involves using two distinct methodologies, namely the term frequency-inverse document frequency and bag of words techniques, to extract relevant attributes from the initial drug review documents. A minimal representation of the original data was extracted by selecting a subset of features. The chosen subset was then used as input to classification algorithms, enhancing the efficiency and accuracy of sentiment analysis. A word embedding model is a sophisticated tool that translates words to vectors, allowing semantic similarities between words to be captured. Bengio et al. [21] pioneered this approach, which has been widely employed in NLP tasks. Carrillo et al. [22] exploited the possibility of using word embeddings to classify patient-generated content based on its polarity (i.e., positive, negative, neutral). Garg S. proposed a medicine recommender system based on the sentiments of the patient reviews that could suggest the top drug for a given disease.

Bow, TF-IDF, Word2Vec, and Manual Feature Analysis were used for vectorization [23]. The field of health-related sentiment analysis has witnessed tremendous advancement with the emergence of deep-learning models in recent years [24]. Yadav et al. [25] identified multiple forms of medical sentiments that can be inferred from users' medical conditions, treatment, and medication. They conducted a study in which they compared the performance of a Convolutional Neural Network (CNN) with conventional ML models such as Random Forest, Multilayer Perceptron, and SVM. The CNN model outperformed the traditional algorithms, resulting in significant advancements in sentiment analysis for health-related aspects. Vikas Goel et al. [26] analyzed sentiments in multilingual tweets, used Google Translator to translate them into English, and proceeded with the process. This paper focused on improving the accuracy of sentiment analysis models by integrating Improved Word Vectors (IWV). They applied two classification techniques, Naïve Bayes (NB) and Recursive Neural Network (RNN), to classify the data based on the sentiment expressed. Mutinda et al. introduced a novel sentiment classification model named LeBERT, which combines Sentiment Lexicon, N-grams, BERT, and CNN within a unified framework to detect the text's sentiment [27] accurately. The authors in [28] noted that utilizing the pre-trained Word2Vec model, which integrates the concept of Implicit Word Vectors (IWV), notably enhanced the model's performance. In a comparative analysis of deep learning models, the research work [29] has shown that the BERT-based BiLSTM neural network classification model outperformed the Convolutional Neural Networks and Long Short-Term Memory recurrent neural networks. In a separate study, ontologies and sentiments were extracted from social media texts using the XLNet model [30]. This model was implemented to uncover indirect relationships in the data and served as a comprehensive context-aware approach for feature extraction. The classification was conducted using the Bi-LSTM model, and the effectiveness of this approach was validated across six drug-related datasets, achieving an impressive accuracy of 98% and an F1 score of 96.4%. Some recent works used a word or sentence-embedding-based sentiment analysis approach [31]. Pre-training embeddings like BERT [32] and Graph Embeddings [33] were used for drug review analysis.

3. Methodology

3.1. Development Framework

We use Exploratory Data Analysis (EDA) to gain insights into the dataset. We then initialize the ML and DL models and train them on the dataset. The performance of these models in classifying the reviews as positive and negative is evaluated using the metrics accuracy, precision, recall, and F1 score. The metrics are then compared to arrive at the most suitable model.

3.2. Dataset Description

The drug review dataset from the University of California, Irvine Machine Learning Repository contains user reviews on various drugs [25]. The dataset encompasses 161,297 entries of patient feedback on specific medications and their associated conditions, each uniquely identified by a unique ID. The features of the dataset include 'drugName' (categorical), representing the name of the drug; 'condition' (categorical), indicating the patient's medical condition; 'review' (text), which contains the patient's detailed review; 'rating' (numerical), providing a 10-star patient rating for the drug; 'date' (date), indicating the review entry date; and 'usefulCount' (numerical), representing the count of users who found the review helpful. The target variable of interest is the sentiment expressed in the review, which is to be predicted. Patients adhere to the prescribed dosage of a specific drug as advised by the healthcare professionals for their medical conditions. They provide reviews while they are actively taking the drug and also upon completing the full prescribed dosage. In these reviews, they may express their opinions on being fully cured, partially cured, not cured at all, or experiencing any side effects due to the medication. The degree of satisfaction is quantified through ratings ranging from 1 to 10.

3.3. Data Pre-processing

The data set is sorted using the distinct medication IDs (data points). The data points with null values for any of the specified features were removed. The training and testing datasets were merged to create an extensive and comprehensive dataset, as sentiments have not been provided for either set. In the absence of explicitly expressed emotions, the ratings in the reviews are used to infer sentiment, serving as the target variable for prediction. Leading and trailing whitespaces in the review text were trimmed, and multiple whitespaces were replaced with a single space to ensure data clarity and consistency. The special characters, non-ASCII characters, HTML tags, punctuation, quotes, URLs, and other formatting elements from the text were removed. Regular expressions were used to clean up the reviews, and all reviews were converted to lowercase. English stop words were also eliminated to enhance model accuracy.

Additionally, all review phrases were stemmed using a snowball stemmer to reduce them to their root form, contributing to the cleanliness and effectiveness of the sentiment analysis model. The text data was cleaned up using the above-said pre-processing operations and was stored under a new feature named 'review_clean'. The 'date' feature in the dataset was removed as it does not contribute to predicting the sentiment of the reviews. New features were also engineered in order to train the ML models.

The following features were derived using the cleaned review text: 'word_count' - the number of words in the review; 'unique_word_count' – the number of unique words in the review; 'letter_count' – the number of letters in the review and 'mean_length_of_words' – the average length of words in the review. The features, 'upper_words_count' – the number of upper case words, 'title_words_count' – the number of title case words, and 'stop_words_count' – the number of stop words, were derived from the unprocessed reviews. After pre-processing, the Drug Review Dataset had 159925 entries with 15 features. The target variable for prediction was the reviewer's rating, which was converted into a binary variable by assigning it a value of 1 if the rating is greater than or equal to five or 0 if otherwise.

3.4. Exploratory Data Analysis

We perform EDA on the pre-processed dataset to understand the distribution and the relationships between the features. The distribution of the features and the relationship between them are visualized using various charts. This study focuses on the top 40 medical conditions for which patients have submitted reviews.

Figure 1 shows the number of drugs available for a given medical condition. Due to space constraints, the bar chart only shows the distribution for the top 25 drugs. The bar chart illustrates that there are multiple drugs to cure a medical condition. Figure 2 illustrates the top 20 drugs that can cure several medical conditions. It is evident that a drug can be used to cure multiple medical conditions, and it is important for the drug recommendation system to choose the drug that leads to the best efficacy.

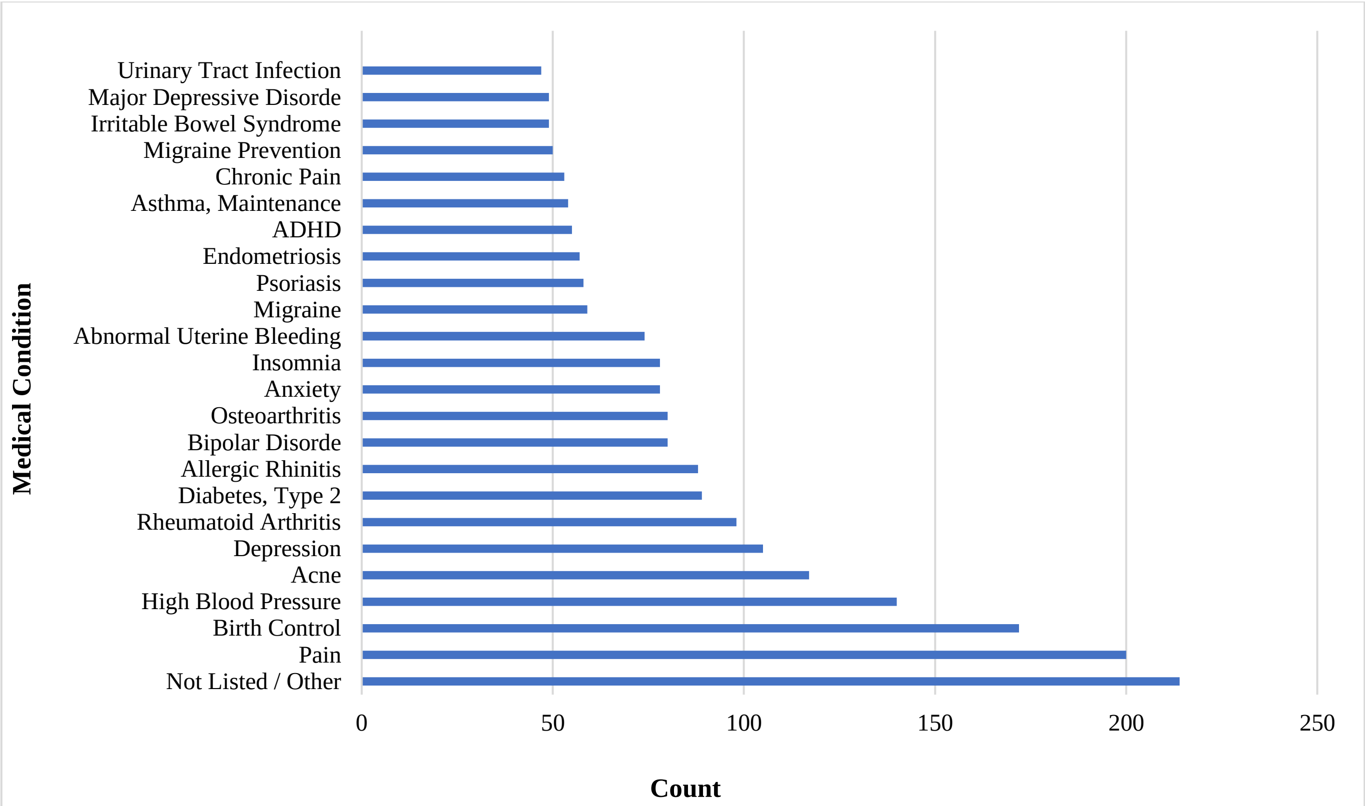

Fig. 1 The bar chart of the number of drugs available for a given medical condition

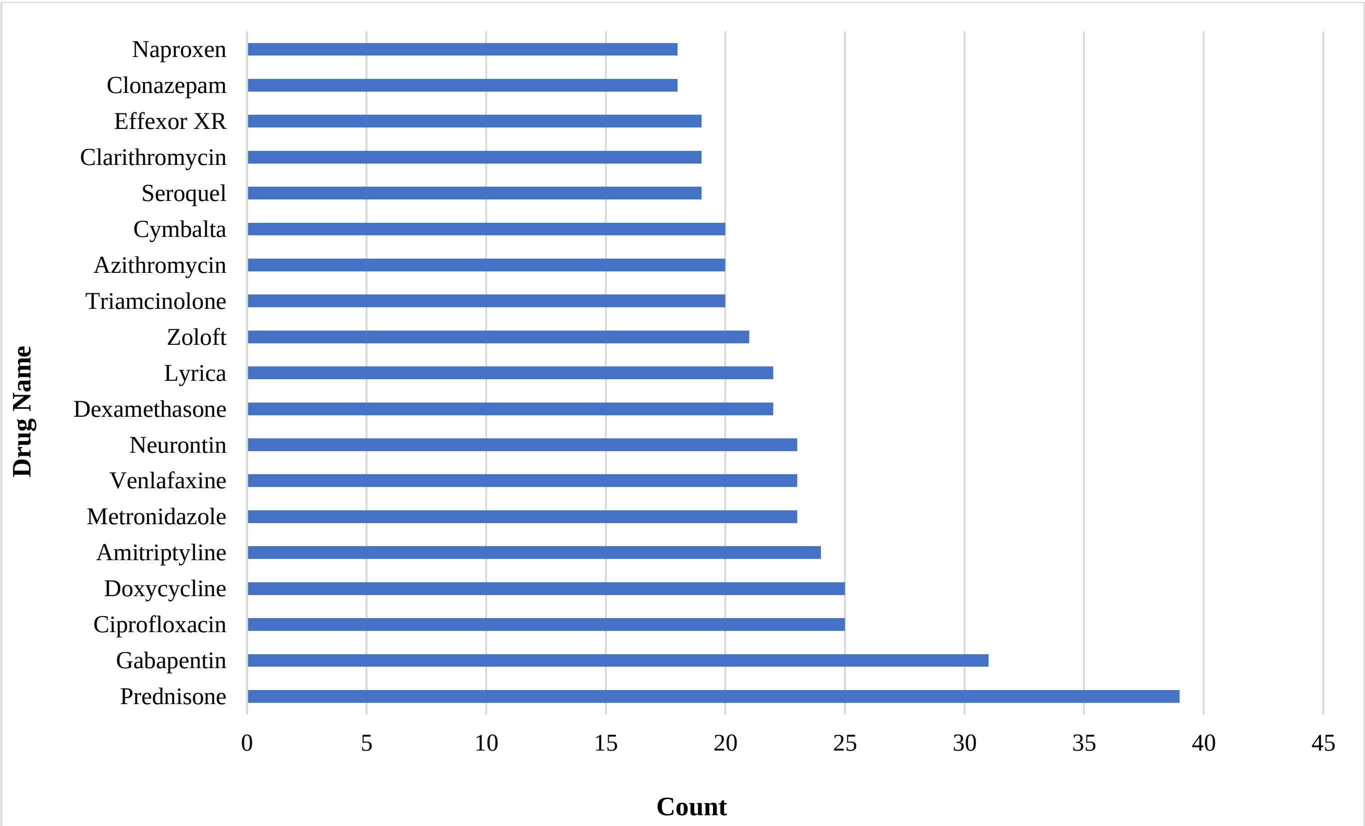

Fig. 2 The bar chart of the number of medical conditions that a drug can cure

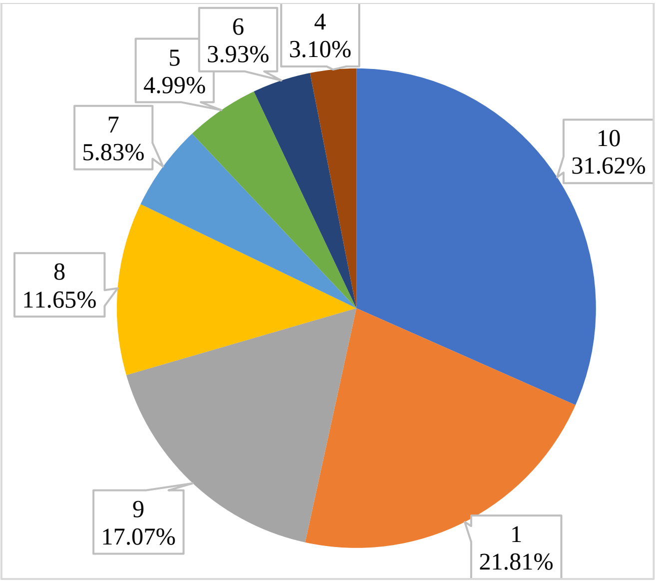

Fig. 3 The Pie chart of the distribution of ratings

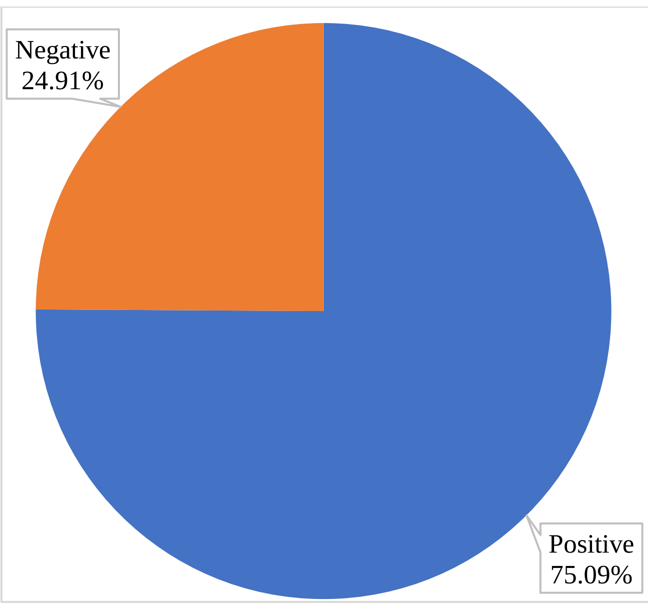

Fig. 4 The Pie chart of the number of positive and negative reviews

Figure 3 shows the distribution of ratings given by the patients based on the drug's impact on their health conditions on a scale of 10. The chart shows that 31.62% of the patients have given a 10 on 10 rating, indicating that they are completely satisfied with the prescribed drug. 21.81% of the reviewers gave 1 out of 10 ratings indicating that the drug they consumed failed to improve their medical condition and may have caused impactful side effects. The other ratings reflect partial satisfaction and dissatisfaction with the drugs. The remaining 46.57% opted for other ratings. We use these ratings to develop the drug recommendation system. Figure 4 shows the percentage of positive and negative reviews in the dataset. One-fourth of the reviews in the dataset are negative, and the remaining are positive.

3.5. Training the ML and DL Models

The review texts were transformed into numeric data using Term Frequency-Inverse Document Frequency (TF-IDF). Equations 1, 2 and 3 give the equations to compute the TF-IDF. The Term Frequency (TF) was computed by dividing the frequency of occurrence of a word in the review by the total number of words in the review. The Inverse Document Frequency (IDF) was computed by taking the logarithm of the result yielded by dividing the total number of reviews in the dataset by the number of reviews in the dataset containing the word. The TF-IDF was computed by multiplying the TF with the IDF.

$$\text{TF} = \frac{\textit{Number of times the word appears in the review}}{\textit{Total number of words in the review}} \tag{1}$$

$$\text{IDF} = \log\left(\frac{\textit{Total number of reviews in the dataset}}{\textit{Number of reviews containing the word}}\right) \tag{2}$$

$$\text{TF-IDF} = \text{TF x IDF} \tag{3}$$

The dataset was divided into training and testing subsets in the ratio of 75:25. The ML models LGBM, Naïve Bayes, LR, and Decision Tree, and the DL models DNN, LSTM, GRU, and BRNN were trained on the training subset, and the performance was analyzed on the testing subset. Below, we briefly describe the salient features of the ML and DL models used in this work.

3.5.1. Multinomial Naive Bayes (MNB)

It is considered a baseline algorithm for any classification task. It is predicated on the idea that word frequency in a review can be used to forecast the likelihood of a patient responding to a particular drug.

3.5.2. Logistic Regression (LR)

It is used to model the relationship between predictor variables and the probability of a positive outcome, which can help predict the likelihood of a positive response to a new drug. It estimates this probability using the features extracted from patient reviews.

3.5.3. Decision Tree (DT)

A decision tree is an ML algorithm that models decisions and their probable outcomes using a tree-like structure. It predicts the output class of a sample using a set of attributes or features represented by a tree's leaf nodes.

3.5.4. Light Gradient-Boosting Machine (LGBM)

It is a gradient-boosting framework that employs decision trees to build an ensemble model. LGBM builds a decision tree leaf by leaf, choosing the leaf with the most significant delta loss, which leads to a more accurate model with shorter training times.

3.5.5. Bidirectional Recurrent Neural Network (BRNN)

RNNs are well-suited for tasks that require a sequential understanding of the data rather than understanding data in isolation. The hidden state in an RNN captures the network's memory of previous inputs and is updated at each time step based on the current input and the previous hidden state. The output of an RNN is dependent both on the current input and the hidden state input. The weights and biases used by RNN are the same for every input processed, enabling the RNN to handle sequences of varied lengths. BRNNs consist of two RNNs processing input data in opposite directions, with their outputs merged to produce the final output.

3.5.6. Long Short-Term Memory (LSTM)

Since RNNs use the same set of weights and biases for all inputs, the gradients vanish or explode as the sequence length becomes too large. Gated architectures like LSTMs are used to alleviate this issue, which controls the flow of information to and from the hidden states.

3.5.7. Gated Recurrent Unit (GRU)

GRU's fundamental framework consists of a recurring hidden layer with gating devices that govern information flow. The gating method enables the model to make selective updates and retrieve information from prior steps, which is particularly helpful for modeling sequential data like time series or natural language sentences. Two gates that manage information flow are found in the recurrent hidden layer of the GRU: an update gate and a reset gate. The update gate chooses how much new data to add to the current hidden state, while the reset gate chooses how much of the prior hidden state to ignore. The GRU learns representations of sequential data that capture the pertinent information for a given task by selectively updating and forgetting information.

3.5.8. Deep Neural Networks (DNN)

The DNN architecture consists of the input layer, hidden dense layers, and a final output layer. The output layer generates the probability distributions for the sentiment categories.

4. Results and Discussions

The performances of various ML and DL models were evaluated using accuracy, precision, recall, and F1 score. The equations to calculate the metrics as mentioned above are given in Equations 4-7. Accuracy is computed by dividing the number of correctly predicted predictions by the total number of predictions. Precision is calculated by dividing the number of true positive predictions by the total number of positive predictions. The recall is computed by dividing the number of true positive predictions by the total actual positives, and the F1 score is the harmonic mean of precision and recall.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP+TN} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (6)$$

$$\text{F1 Score} = \frac{2 \; X \; Precision \; X \; Recall}{Precision+Recall} \quad (7)$$

Tables 1 and 2 show the performance metrics of the ML and DL models chosen for this study. Among the ML models, the Light Gradient-Boosting Machine performs the best, as indicated by the highest F1 score.

While having a comparatively low accuracy, Logistic Regression excels in the recall metric, suggesting that it performs well in identifying positive instances but may also generate more false positives.

Multinominal Naïve Bayes lags in terms of recall and F1 score, possibly due to its simplistic assumption of feature independence despite having good precision. Decision Tree has performed the worst among the ML models taken up for the study.

Table 1. Performance of ML models

Models	Accuracy	Precision	Recall	F1 score
MNB	0.76	0.80	0.76	0.78
LR	0.72	0.71	0.83	0.76
DT	0.70	0.68	0.64	0.66
LGBM	0.81	0.81	0.81	0.81

Table 2. Performance of DL models

Models	Accuracy	Precision	Recall	F1 score
Bi-RNN	0.93	0.91	0.89	0.90
LSTM	0.87	0.88	0.87	0.88
GRU	0.88	0.87	0.87	0.87
DNN	0.82	0.8	0.82	0.81

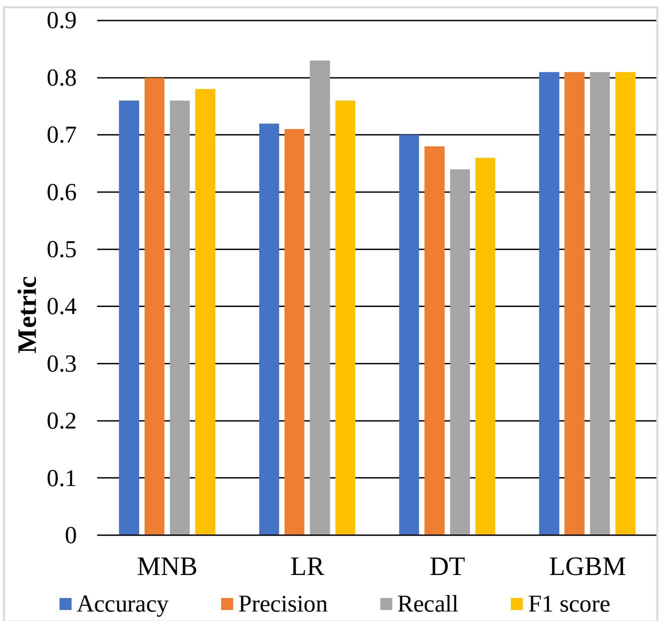

Fig. 5 Performance comparison of ML models

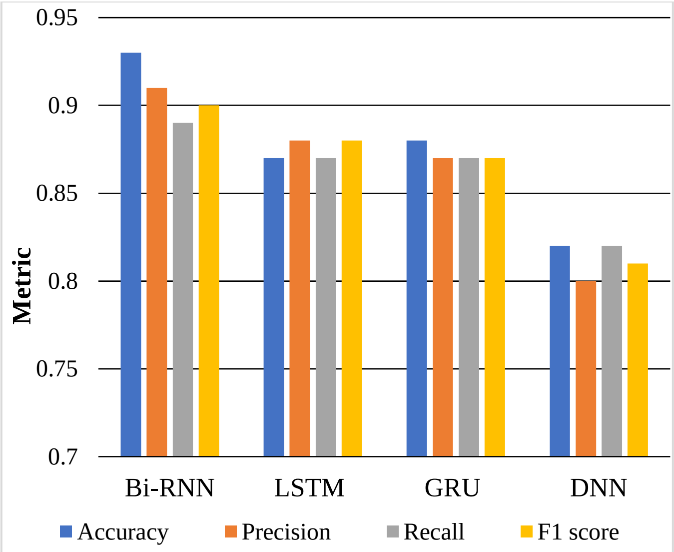

Fig. 6 Performance comparison of DL models

	condition	drugName	mean_total_pred
	Allergic Rhinitis	Singulair	0.261661
	Allergic Rhinitis	Montelukast	0.160027
	Allergic Rhinitis	Olopatadine	0.147487
	Allergic Rhinitis	Allegra ODT	0.136952
	Allergic Rhinitis	Flonase	0.130095
	Allergic Rhinitis	Allegra	0.121752
	Allergic Rhinitis	Astepro	0.116896
	Allergic Rhinitis	Zyrtec-D	0.115883
	Allergic Rhinitis	Fluticasone	0.110171
	Allergic Rhinitis	Chlor-Trimeton	0.109647

Fig. 7 Drugs for allergic rhinitis

Among the DL models, it is evident that the Bidirectional RNN stands out as the top-performing model with the highest F1 score of 90%. Long Short-Term Memory is the next best performing model with an F1 score of 88%. Deep Neural Networks shows an accuracy of 82% but falls behind in precision, potentially indicating a higher rate of false positives. The Gated Recurrent Unit exhibits a slightly lower F1 score compared to Bi-RNN and LSTM. Figures 5 and 6 visually show the comparison of the performance of ML and DL models. To determine the most suitable drug for a given medical condition, we multiply the useful count of the review with the prediction of the sentiment analysis models. Figure 7 displays the snapshot of drugs recommended for the medical condition 'Allergic Rhinitis'. The drug with the highest predicted score is recommended for this condition.

5. Conclusion

In conclusion, we have developed a drug recommendation system using the drug reviews from the UCI Drug Review Dataset. The performance of ML (Light Gradient-Boosting Machine, Multinomial Naïve Bayes, Logistic Regression, and Decision Tree) and DL (Bidirectional Recurrent Neural Network, Long Short-Term Memory, Gated Recurrent Unit, and Deep Neural Networks) models were analyzed in predicting the sentiment of the reviews. Among the ML and DL models, LGBM and Bi-RNN performed the best, respectively. Based on the predicted sentiments, we then predict the most suitable drug for a given medical condition. Overall, this research has demonstrated a robust approach to extracting opinions from drug reviews in social media forums through the application of NLP techniques integrated with ML and DL models. Additionally, integrating additional data sources, such as clinical trials and customer feedback, could provide a more comprehensive understanding of drug safety and efficacy. Refining these methodologies allows us to advance toward more personalized and effective healthcare interventions based on patient feedback and empirical data.

References

[1] Pantea Keikhosrokiani, Katheeravan Balasubramaniam, and Minna Isomursu, "Drug Recommendation System for Healthcare Professionals' Decision-Making Using Opinion Mining and Machine Learning," *Digital Health and Wireless Solutions, Communications in Computer and Information Science*, vol. 2084, pp. 222-241, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] Hariprasad Sampathkumar, Xue-wen Chen, and Bo Luo, "Mining Adverse Drug Reactions from Online Healthcare Forums Using Hidden Markov Model," *BMC Medical Informatics and Decision Making*, vol. 14, pp. 1-18, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[3] Felix Gräßer et al., "Aspect-Based Sentiment Analysis of Drug Reviews Applying Cross Domain and Cross-Data Learning," *Proceedings of the 2018 International Conference on Digital Health*, New York, NY, USA, pp. 121-125, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[4] Ioannis Korkontzelos et al., "Analysis of the Effect of Sentiment Analysis on Extracting Adverse Drug Reactions from Tweets and Forum Posts," *Journal of Biomedical Informatics*, vol. 62, pp. 148-158, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[5] Shuroug A. Alowais et al., "Revolutionizing Healthcare: The Role of Artificial Intelligence in Clinical Practice," *BMC Medical Education*, vol. 23, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[6] Mayur Wankhade, Annavarapu Chandra Sekhara Rao, and Chaitanya Kulkarni, “A Survey on Sentiment Analysis Methods, Applications, and Challenges,” *Artificial Intelligence Review*, vol. 55, pp. 5731-5780, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Salud María Jiménez-Zafra et al., “How Do We Talk About Doctors and Drugs? Sentiment Analysis in Forums Expressing Opinions for Medical Domain,” *Artificial Intelligence in Medicine*, vol. 93, pp. 50-57, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[8] Hamed Taherdoost, and Mitra Madanchian, “Artificial Intelligence and Sentiment Analysis: A Review in Competitive Research,” *Computers*, vol. 12, no. 2, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Nirmal Varghese Babu, and E. Grace Mary Kanaga, “Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review,” *SN Computer Science*, vol. 3, pp. 1-20, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] Sunir Gohil, Sabine Vuik, and Ara Darzi, “Sentiment Analysis of Health Care Tweets: Review of the Methods Used,” *JMIR Public Health and Surveillance*, vol. 4, no. 2, pp. 1-10, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[11] Anastazia Zunic, Padraig Corcoran, and Irena Spasic, “Sentiment Analysis in Health and Well-Being: Systematic Review,” *JMIR Medical Informatics*, vol. 8, no. 1, pp. 1-22, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] Rezaul Haque et al., “Data-Driven Solution to Identify Sentiments from Online Drug Reviews,” *Computers*, vol. 12, no. 4, pp. 1-22, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Jin-Cheon Na et al., “Sentiment Classification of Drug Reviews Using a Rule-Based Linguistic Approach,” *The Outreach of Digital Libraries: A Globalized Resource Network*, pp. 189-198, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[14] Lorraine Goeuriot et al., “Sentiment Lexicons for Health-Related Opinion Mining,” *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, Miami Florida USA, pp. 219-226, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[15] Rezaul Haque et al., “Multi-Class Sentiment Classification on Bengali Social Media Comments Using Machine Learning,” *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 21-35, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[16] Rezaul Haque et al., “A Comparative Analysis on Suicidal Ideation Detection Using NLP, Machine, and Deep Learning,” *Technologies*, vol. 10, no. 3, pp. 1-15, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] Victoria Bobicev et al., “Learning Sentiments from Tweets with Personal Health Information,” *Advances in Artificial Intelligence, Canadian AI 2012, Lecture Notes in Computer Science*, vol. 7310, pp. 37-48, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[18] Sairamvinay Vijayaraghavan, and Debraj Basu, “Sentiment Analysis in Drug Reviews Using Supervised Machine Learning Algorithms,” *Arxiv*, pp. 1-9, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[19] Nadhem Zmandar, Mahmoud El-Haj, and Paul Rayson, “Multilingual Financial Word Embeddings for Arabic, English and French,” *2021 IEEE International Conference on Big Data*, Orlando, FL, USA, pp. 4584-4589, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[20] Tianhua Chen et al., “Sentiment Classification of Drug Reviews Using Fuzzy-rough Feature Selection,” *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, New Orleans, LA, USA, pp. 1-6, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[21] Yoshua Bengio, Réjean Ducharme, and Pascal Vincent, “A Neural Probabilistic Language Model,” *Advances in Neural Information Processing Systems*. pp. 1137-1155, 2003. [CrossRef] [Google Scholar] [Publisher Link]

[22] Jorge Carrillo-de-Albornoz, Javier Rodríguez Vidal, and Laura Plaza, “Feature Engineering for Sentiment Analysis in E-Health Forums,” *PLoS One*, vol. 13, no. 11, pp. 1-25, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[23] Satvik 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)*, Noida, India, pp. 175-181, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[24] Cristóbal Colón-Ruiz, and Isabel Segura-Bedmar, “Comparing Deep Learning Architectures for Sentiment Analysis on Drug Reviews,” *Journal of Biomedical Informatics*, vol. 110, pp. 1-11, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[25] Shweta Yadav et al., “Medical Sentiment Analysis Using Social Media: Towards building a Patient Assisted System,” *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, pp. 1-8, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[26] Vikas Goel, Amit Kr. Gupta, and Narendra Kumar, “Sentiment Analysis of Multilingual Twitter Data using Natural Language Processing,” *2018 8th International Conference on Communication Systems and Network Technologies (CSNT)*, Bhopal, India, pp. 208-212, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[27] James Mutinda, Waweru Mwangi, and George Okeyo, “Sentiment Analysis of Text Reviews Using Lexicon-Enhanced Bert Embedding (LeBERT) Model with Convolutional Neural Network,” *Applied Sciences*, vol. 13, no. 3, pp. 1-14, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[28] Seyed Mahdi Rezaeinia et al., “Sentiment Analysis Based on Improved Pre-Trained Word Embeddings,” *Expert Systems with Applications*, vol. 117, pp. 139-147, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[29] Asmaa Hashem Sweidan, Nashwa El-Bendary, and Haytham Al-Feel, "Sentence-Level Aspect-Based Sentiment Analysis for Classifying Adverse Drug Reactions (ADRs) Using Hybrid Ontology-XLNet Transfer Learning," *IEEE Access*, vol. 9, pp. 90828-90846, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[30] Shila Sumol Jawale, and S.D. Sawarker, "Amalgamation of Embeddings with Model Explainability for Sentiment Analysis," *International Journal of Applied Evolutionary Computation*, vol. 13, no. 1, pp. 1-24, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[31] Mehmet Bozuyla et al., "Sentiment Analysis of Turkish Drug Reviews with Bidirectional Encoder Representations from Transformers," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 23, no. 1, pp. 1-17, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[32] Xinkun Hao et al., "Enhancing Drug–Drug Interaction Prediction by Three-Way Decision and Knowledge Graph Embedding," *Granular Computing*, vol. 8, pp. 67-76, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[33] Surya Kallumadi, and Felix Grer, "*Drug Reviews (Drugs.com)*," UC Irvine Machine Learning Repository. 2018. [CrossRef] [Publisher Link]