RESEARCH ARTICLE

INTERNATIONAL RESEARCH JOURNAL OF MULTIDISCIPLINARY TECHNOVATION

Asian Research Association

DOI: 10.54392/irjmt26510

Reliable Drug Recommendations using a Hybrid CNN-BiLSTM and Whale Algorithm-Optimized Aspect-Based Sentiments

Suruchi Chawla [a], Aakanksha [a,*], Madhu Rani [a]

[a] Department of Computer Science, Shaheed Rajguru College of Applied Sciences for Women, University of Delhi, Delhi, India

* Corresponding Author Email: aakanksha.1@rajguru.du.ac.in

DOI: https://doi.org/10.54392/irjmt26510

Received: 24-04-2026; Revised: 22-08-2026; Accepted: 10-09-2026; Published: 15-09-2026

Abstract: The growing availability of drug reviews from patients has paved the way for future patient-centric and data-enabled healthcare support; however, the majority of existing sentiment-based drug recommendation systems use overall sentiment (sentiment polarity) to generate drug recommendations. In this study, an aspect-aware and optimized deep-learning-based drug recommendation system is proposed using sentiment analysis. A combined Convolutional Neural Network and Bidirectional Long Short-Term Memory architecture is employed to leverage important local semantics as well as contextual and sequential dependencies from drug reviews. The model is extended into an explainable drug recommendation system that maps user-reported symptoms to patient-opinion-informed drug rankings. Sentiments are analyzed at the aspect level, focusing on effectiveness, side effects, dosage, safety, and cost. To achieve better results, the Whale Optimization Algorithm is used to assign optimal weights to each aspect, thereby ranking drugs according to data-driven aspect preferences learned from patient reviews. Experimental evaluation on the drug dataset, comprising patient reviews, demonstrates that the proposed hybrid sentiment classification model achieves 99.80% accuracy. It outperforms traditional machine learning and deep learning approaches. The proposed framework not only improves sentiment classification accuracy but also provides transparent and improved top-k drug recommendations, offering explainable patient-opinion-informed recommendation rankings derived from large-scale review data. The generated rankings are derived from patient-generated review data and should therefore be interpreted as opinion-aware recommendations rather than clinically validated treatment recommendations.

Keywords: Sentiment Analysis, Convolutional Neural Network (CNN)-BiLSTM, Deep Learning (DL), Drug Recommendation System (DRS), Whale Optimization Algorithm (WOA), Aspect-Based Sentiment Analysis (ABSA).

1. Introduction

The healthcare sector has experienced major changes driven by digital technology over the past decade, and these changes can be seen across multiple domains. It is evident that a digital transformation is underway in health systems, and digital technologies are being utilized to develop innovative solutions for better healthcare services and improved patient-opinion-informed drug recommendations for specific medical conditions [1, 2]. As healthcare continues to digitize, data from online drug review platforms has become a valuable resource [3]. This data includes details about how well a medication worked, its side effects, and patient opinions, and it offers unparalleled insight into real-world evidence. Such data can be used to build sophisticated decision-support tools that improve the quality of care, offer better treatment options, and improve patient outcomes [4]. The challenge is that patient reviews are generally written in free-form text in a colloquial style that mixes clinical terminology with colloquial terms, slang, abbreviations, and opinions [5]. All of this makes it difficult to extract meaningful information using traditional methods. To overcome these difficulties, researchers have started turning to NLP (Natural Language Processing) techniques to evaluate these reviews and transform raw text into useful information [6-8]. This diversity makes it difficult to extract useful insights with traditional computational methods.

Sentiment analysis of drug reviews is a promising technique for facilitating data-driven drug recommendation and decision-making in the medical and healthcare domains [9]. Most existing studies on drug reviews are limited to coarse-grained sentiment polarity classification, that is, positive, negative, and neutral, which cannot capture subtle sentiment information about the use of a drug; for instance, a drug may be effective but have serious side effects or a high price. Therefore, Aspect-based Sentiment Analysis

(ABSA) for drug reviews has attracted growing attention from researchers. In ABSA, sentiment analysis is performed at a more fine-grained level, and the relevant aspects include effectiveness, side effects, dosage, safety, and price. Deep learning architectures such as LSTM, CNN, and BiLSTM networks have been more effective in performing aspect-level sentiment analysis of drug reviews [10]. This is due to their ability to model contextual and sequential dependencies in medical texts. A hybrid CNN-BiLSTM model combines local feature learning with long-range semantic encoding and is therefore an ideal model for understanding and analyzing complex healthcare-related texts.

Despite these advances, two critical challenges remain: (i) existing ABSA models assign equal weight to each aspect, while in real-world clinical settings some aspects are more important than others [11], and (ii) deep-learning-based Recommender Systems (RSs) are not interpretable, which is a critical concern in healthcare applications. To address these limitations, an aspect-based sentiment-driven drug recommendation framework is proposed. In this research, a hybrid model is developed using CNN and BiLSTM for aspect extraction and sentiment analysis, while the Whale Optimization Algorithm is employed to optimize the weights of different sentiment aspects for clinically relevant drug ranking. Furthermore, an explainability module provides aspect-wise contribution scores, which help users understand the reasoning behind a specific recommendation. By utilizing large-scale patient reviews, the framework generates drug recommendations that can be employed in real-world healthcare decision-support systems.

The rest of the paper is structured as follows: Section 2 reviews the literature on drug recommendations based on sentiment analysis and aspect-level analysis. Section 3 presents the methodology, including the model formulation, evaluation metrics, and experimental setup. Section 4 presents the experimental results and comparative analyses. Section 5 discusses the findings, limitations, and implications of the study. Finally, Section 6 concludes the paper with a summary of the work and its future scope.

2. Related Work

In [12], Srishailam et al. proposed a patient-centered drug recommendation framework that used machine learning by considering customer reviews. They showed how recommendation quality can be improved by extracting sentiment information and supplementing approaches based solely on structured information. Nevertheless, their work performs sentiment analysis only at the review level using an existing framework and therefore cannot leverage aspect-level clinical sentiments for individual drug ranking.

Sequeira et al. [13] proposed “an intelligent disease prediction and DRS using ML-based supervised learning techniques (decision tree, support vector machine).” Although the authors report decent performance in disease prediction and drug recommendation, their model considers only structured patient data and not textual patient opinions.

In [14], Wang et al. proposed a deep-reinforcement-learning-based model for medication recommendation in a clinical setting that recommends optimal antibiotic dosages. However, although the model was tested on real clinical datasets, it was limited to structured clinical data and did not consider sentiment or patient-reported reviews as a way of tailoring recommendations to the patient’s subjective experience.

Al-Hadhrami et al. [9] employed “CNN- and LSTM-based DL models for sentiment analysis of drug reviews and compared the performance with conventional ML models.” The results showed substantial improvements in classification performance. However, the emphasis was placed solely on binary and multi-class classification of sentiment polarities. No attention was given to identifying the multi-aspect sentiment features that play a pivotal role in decision-making tasks.

Wang et al. [15] addressed the multi-aspect sentiment analysis problem in the health domain, where several drug aspects (e.g., efficacy and side effects) are identified to train a deep neural network model for sentiment prediction. Although they showed that aspect-level sentiment prediction is essential for medical treatment, their study did not differentiate the priority among aspects; in practice, different aspects have different levels of importance in drug selection.

Lin et al. [16] recently introduced a neural attention-based aspect sentiment model for medical reviews. The model showed higher interpretability and achieved better performance than related studies. However, it was not trained with a mechanism to optimize weights among aspects, and no recommendation model was employed for the predicted sentiments.

Martino et al. [17] proposed explainable deep learning models for medical text analysis, focusing on the explainability of healthcare AI models. However, they did not focus on the drug recommendation problem and did not consider aspect-based sentiment for improving ranking.

Kumar *et* al. [11] employed WOA to optimize the hyperparameters of a DL model for medical text classification and showed its fast convergence and higher classification accuracy. However, the optimization was carried out only for model parameters, without learning the relative importance of aspects.

He et al. [18] tackled sentiment analysis by focusing on the sentiment polarity of comments. They proposed a hybrid BERT-CNN-BiLSTM model. In this model, BERT handles word embeddings, CNN extracts features, and BiLSTM captures long-term dependencies before making the final prediction. Their approach boosted accuracy by as much as 27.78% compared with BERT, CNN, BiLSTM, and even BERT-BiLSTM setups.

Li et al. [19] developed an ensemble DL model consisting of CNN and BiLSTM for healthcare sentiment analysis, which outperformed state-of-the-art approaches in modeling both short- and long-range dependencies in patient text data. However, they focused only on improving sentiment classification performance without addressing healthcare recommendation ranking or explainability.

Vanam [20] advocated the use of interpretable and sentiment-informed drug recommender systems to overcome the limitations of black-box deep learning models and thus highlighted the need for aspect-level sentiment analysis in conjunction with optimization and interpretability when designing clinically validated and interpretable recommendation models. The comparative analysis of existing drug recommendation and sentiment analysis approaches is given in Table 1.

Table 1. Comparative analysis of existing drug recommendation and sentiment analysis approaches

Author (Year)	Aim	Methods	Advantages	Limitations
Satvik *et al.* (2024) [21]	To improve drug recommendation using patient sentiment.	Machine learning classifiers with sentiment features extracted from patient reviews	Demonstrated that sentiment information improves recommendation accuracy	Restricted to coarse-grained (binary) sentiment; lacks aspect-level insight
Nayak *et al*. (2023) [22]	To predict diseases and recommend drugs using clinical attributes	Decision tree and SVM - Supervised learning algorithms	Effective use of structured medical data	Patient opinions and textual feedback are not considered
Wang *et al.* (2024) [14]	To recommend optimal antibiotic dosages in critical care	Deep reinforcement learning on clinical records	Clinically validated and data-driven dosage optimization	Does not incorporate patient experience or review-based information
Al-Hadhrami *et al.* (2024) [9]	To classify sentiment in patient drug reviews	CNN and LSTM architectures	Superior performance compared to traditional ML models	Focus limited to sentiment classification without recommendation
Wang et al. (2021) [15]	To analyze sentiment across multiple drug aspects	DNN for aspect-based sentiment analysis	Highlights importance of aspect-level sentiment in healthcare	All aspects treated uniformly without prioritization
Lin *et al.* (2023) [16]	To improve interpretability of medical sentiment models	Attention-based neural aspect sentiment model	Better transparency and performance	Aspect importance not optimized; no recommendation stage
Martino *et al.* (2023) [17]	To improve explainability in medical text analytics	Explainable deep learning frameworks	Enhances trust in healthcare AI systems	Not focused on drug recommendation tasks
Kumar *et al.* (2025) [11]	To advocate interpretable sentiment-aware drug recommendation	Conceptual sentiment-driven recommendation framework	Emphasizes explainability and clinical relevance	Lacks implementation and experimental validation
He *et al.* (2023) [18]	To improve sentiment analysis accuracy by determining the sentiment polarity of comments.	Hybrid BERT-CNN-BiLSTM model	Achieved higher accuracy than BERT, CNN, BiLSTM, and BERT-BiLSTM models, improving performance by up to 27.78%.	Computationally intensive due to multiple DL components and requires large training data

3. Methodology

3.1 CNN-BiLSTM Model

A hybrid CNN-BiLSTM model is designed to capture local sentiment-oriented information and long-range contextual information in medical review text [19]. A CNN is used to extract n-gram features, and a BiLSTM network is used to model the forward and backward sequential dependencies of patient opinions [8, 23, 24].

3.1.1 Model Description

Assume that a preprocessed review is a token sequence

$$X = \{w_1, w_2, \dots, w_T\} \quad (1)$$

That has been computed after padding or truncating [8]. Here, T denotes the length of the input sequence. The input sequence is then fed into an embedding layer to obtain the embedding vector for each token, and a corresponding embedding matrix is generated as follows

$$E = [\mathrm{e}_1, \mathrm{e}_2, \dots, \mathrm{e}_T], \mathrm{e}_i \in \mathbb{R}^d \quad (2)$$

Here, d denotes the dimension of the embedding vectors.

3.1.2 Convolutional Feature Extraction

A one-dimensional convolution operation is performed over the embedding matrix to extract local sentiment-carrying sequences. For a given convolution filter, where the kernel size, the convolution result at position is i is given by:

$$c_i = f(\mathrm{W}_c \cdot E_{i:i+k-1} + b_c) \quad (3)$$

Where the bias is term, and denotes a nonlinear activation function (ReLU). Multiple filters are employed to capture a variety of local features related to the sentiments of interest (e.g., efficacy, side effects, and side reactions).

3.1.3 Max-Pooling Layer

For each feature map, a max-pooling operation retains only the maximum value:

$$\hat{c} = \max\{c_1, c_2, \dots, c_{T-k+1}\} \quad (4)$$

This operation is performed to retain only the most useful information (i.e., the most important sentiment signals) from each feature map.

3.1.4 Bidirectional LSTM Layer

Finally, the features extracted by the CNN are fed into a Bi-LSTM to learn context-dependent representations. The Bi-LSTM consists of a forward LSTM that reads the input sequence from 1 to T and a backward LSTM that reads the sequence from T to 1 [25]. The hidden states of the forward and backward LSTMs at time step t are computed as given in Eq. (5):

$$\overrightarrow{h_t} = \mathrm{LSTM}_f(\hat{c}_t), \overleftarrow{h_t} = \mathrm{LSTM}_b(\hat{c}_t) \quad (5)$$

Finally, the hidden state at time step t is the concatenation of the forward and backward hidden states [24]:

$$h_t = [\overrightarrow{h_t} \parallel \overleftarrow{h_t}]$$

Bi-LSTM is employed because it can capture sentiment relationships that depend on the context both before and after a concept, which is typical in clinical text.

3.1.5 Output Layer and Sentiment Prediction

The final hidden state is used to train a fully connected network with a Softmax output function to predict the probability distribution of sentiment labels:

$$\hat{y} = \mathrm{Softmax}(W_o h + b_o) \quad (6)$$

Where W_o and b_o are trainable parameters. The Softmax function is defined as:

$$\mathrm{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \quad (7)$$

with denoting the number of sentiment labels.

3.1.6 Model Training

The network is trained using categorical cross-entropy loss:

$$\mathcal{L} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \quad (8)$$

Where and represent the actual and predicted sentiment labels, respectively [26]. To avoid overfitting, the AdamW optimizer is used, with the learning rate adapted for each parameter to achieve model convergence. As shown in Figure 1, the convolutional and BiLSTM layers are jointly trained to learn sentence representations for extracting complex features, thereby improving sentiment classification performance. The sentiment classification model is a key component of the system because it provides the predicted sentiment labels for aspect-level sentiment analysis, optimization, and explainable drug recommendation [27].

3.2 Aspect Extraction

The rule-based strategy is employed for the sake of interpretability [28]. Suppose the review comprises a sequence of tokens. For each predefined aspect, where:

$$A = \{\text{effectiveness, side effects, dosage, safety, cost}\}$$

A review is associated with an aspect if:

$\exists\, w_i \in R$ s.t. $w_i \in K_{a_j}$

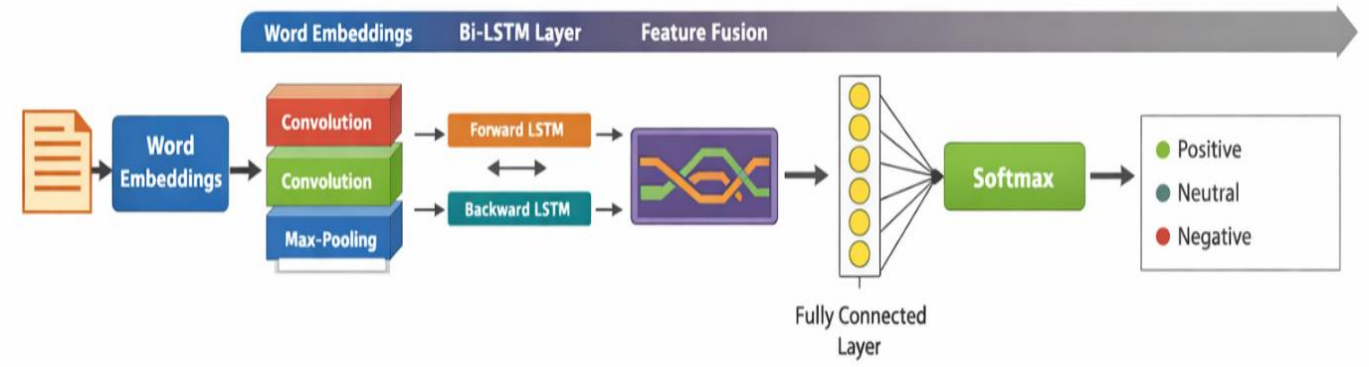

Figure 1. Design of CNN and BiLSTM for drug sentiment classification.

Table 2. Aspect Definitions Used for Fine-Grained Sentiment Analysis

Aspect	Description	Example Keywords
Effectiveness	Reflects the perceived ability of a drug to treat or control the medical condition	effective, works, relief, improvement
Side Effects	Captures adverse reactions or unwanted effects experienced by patients	nausea, headache, fatigue, dizziness
Dosage	Relates to dosage amount, frequency, or administration concerns	dose, dosage, daily, mg, timing
Safety	Indicates overall safety, long-term use concerns, or risk perception	safe, risky, harm, dependency
Cost	Represents affordability and economic burden of the medication	expensive, cost, price, affordable

Where denotes the keyword dictionary corresponding to the aspect.

This process allows a single review to be linked with multiple aspects, reflecting the multifaceted nature of patient feedback [29]. A snippet of the extracted aspects and their descriptions is summarized in Table 2.

3.3 Aspect-Level Sentiment Assignment

For each extracted aspect instance, the sentiment polarity predicted by the CNN-BiLSTM classifier is assigned directly. Let denote the sentiment label of review, corresponding to negative, neutral, and positive sentiments, respectively [8, 30].

The sentiment score of aspect a_j for drug d is computed as:

$$S_{d,a_j} = \frac{1}{N_{d,a_j}} \sum_{r=1}^{N_{d,a_j}} s_r \quad (9)$$

Where N_{d,a_j} denotes the number of reviews for a drug d that mention the aspect a_j.

This aggregation produces aspect-wise sentiment profiles for each drug, which are later used for optimized weighting and ranking.

3.4 Whale Optimization-Based Aspect Weighting

Since different aspects contribute unequally to drug selection, WOA is employed to learn optimal aspect weights [31]. Each whale represents a candidate weight vector as given in Eq. (10):

$$\mathrm{W} = [w_1, w_2, \ldots, w_m], \sum_{j=1}^{m} w_j = 1 \quad (10)$$

Where m is the number of aspects

The fitness function for a candidate solution is defined as:

$$\mathcal{F}(\mathrm{W}) = \sum_{j=1}^{m} w_j \cdot S_{d,a_j} \quad (11)$$

WOA updates whale positions using encircling and spiral-movement strategies:

$$\mathrm{X}(t+1) = \mathrm{X}^*(t) - \mathrm{A} \cdot | \mathrm{C} \cdot \mathrm{X}^*(t) - \mathrm{X}(t \quad (12)$$

Where is the best solution at iteration, and are coefficient vectors controlling exploration and exploitation.

The optimized weight vector obtained after convergence reflects data-driven aspect priorities learned from patient-generated review sentiment.

3.5 Trust Score Computation

To enhance the robustness of review-based ranking, the trust score for a drug is defined as follows:

$$T_d = \alpha \cdot \mu_d + \beta \cdot \log(N_d) - \gamma \cdot \sigma_d \quad (13)$$

Where:

- $\mu_d$is the mean weighted aspect sentiment score,
- $N_d$is the number of reviews,
- $\sigma_d$is the standard deviation of sentiment scores,
- $\alpha, \beta, \gamma$are tunable coefficients.

This formulation penalizes drugs with sparse or inconsistent feedback, thereby improving robustness [32].

3.6 Drug Recommendation and Ranking

For a given medical condition, candidate drugs are filtered using condition matching. The final recommendation score for a drug is computed as:

$$R_d = \mathcal{F}(\mathrm{W}) * T_d \quad (14)$$

Drugs are ranked in descending order, producing a prioritized top-k recommendation list. A schematic overview of this ranking pipeline is illustrated in Figure 2.

3.7 Model Evaluation Metrics

The quantitative evaluation metrics used to measure the performance of the CNN-BiLSTM sentiment classification model are given in Table 3. Assume that the actual and predicted labels are denoted accordingly.

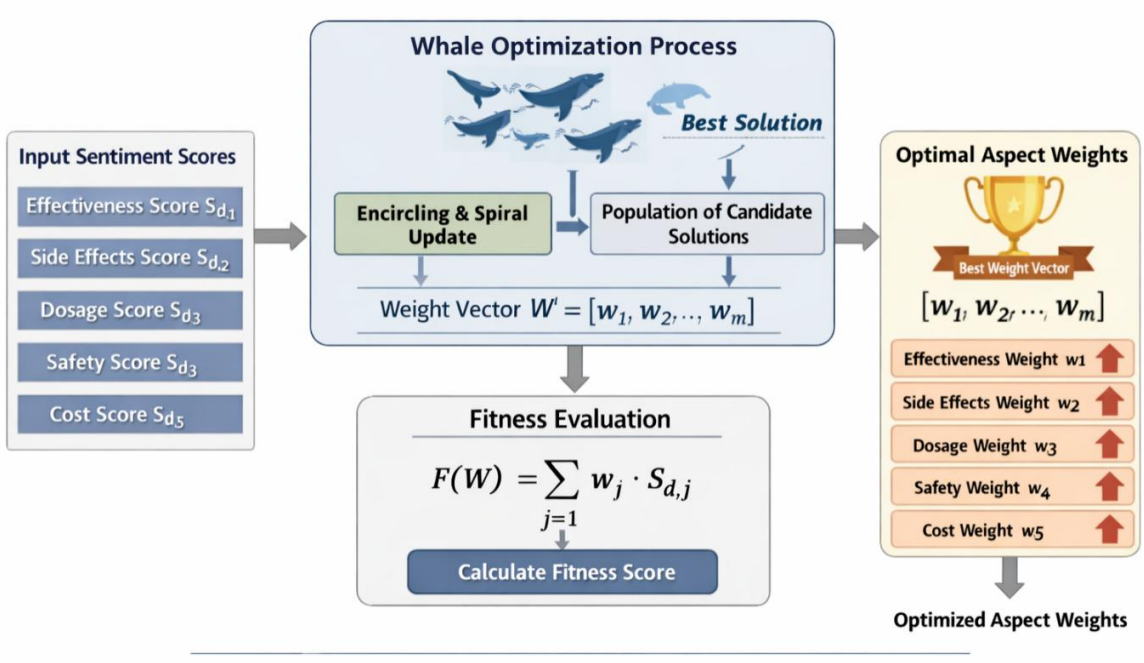

Figure 2. Whale Optimization-Based Aspect Weighting.

Table 3. Performance Metrics for CNN–BiLSTM Sentiment Classification

Metric	Mathematical Formula	Description
Accuracy [33]	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$ (15)	Predicts the overall correctness of the model across all classes.
Precision [6]	$Precision = \frac{TP}{TP + FP}$ (16)	Measures the proportion of predicted positive samples that are actually positive. Indicates model exactness.
Recall [21]	$Recall = \frac{TP}{TP + FN}$ (17)	Measures the ratio of actual positive samples that are correctly identified. Indicates model completeness.
F1-Score [33]	$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$ (18)	Harmonic mean of Precision and Recall.

Where TP , TN , FP , and FN represent true positives, true negatives, false positives, and false negatives, respectivel y.

A confusion matrix is employed to observe class-wise correct and incorrect predictions. Further, Cohen's Kappa coefficient is also used to assess class-wise agreement between actual and predicted class labels beyond chance [33]:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (19)$$

Where p_o is the observed agreement and p_e is the agreement expected by random chance [34]. Cohen's Kappa coefficient is more useful for evaluating class-wise classification performance for multi-class healthcare datasets.

3.7.1 Micro Average

Micro average calculates global metric scores by computing TP, FP, and FN for each class and then calculating the metric based on their sum [35]. Therefore, it treats every prediction equally, regardless of class size, which is useful for evaluating overall system performance [31].

$$\text{Precision}_{micro} = \frac{\sum_i TP_i}{\sum_i (TP_i + FP_i)} \quad (20)$$

$$\text{Recall}_{micro} = \frac{\sum_i TP_i}{\sum_i (TP_i + FN_i)} \quad (21)$$

$$F1_{micro} = \frac{2 \cdot \text{Precision}_{micro} \cdot \text{Recall}_{micro}}{\text{Precision}_{micro} + \text{Recall}_{micro}} \quad (22)$$

3.7.2 Weighted Average

"Weighted average calculates the metric for each class and computes the weighted average of the metrics, where the weights are the number of true instances (support) in each class" [35]. Therefore, the class with more instances dominates the score, making the metric more reflective of the actual data distribution [36].

$$\text{Precision}_{weighted} = \frac{\sum_i (n_i \cdot \text{Precision}_i)}{\sum_i n_i} \quad (23)$$

$$\text{Recall}_{weighted} = \frac{\sum_i (n_i \cdot \text{Recall}_i)}{\sum_i n_i} \quad (24)$$

$$F1_{weighted} = \frac{\sum_i (n_i \cdot F1_i)}{\sum_i n_i} \quad (25)$$

Where n_i= equals the number of true instances (support) for a class$i$.

The evaluation protocol adopted in this study is summarized in Table 4. Multiple quantitative measures were employed to assess classification performance, comparative effectiveness, and recommendation quality.

3.8 Experimental Setup

The proposed framework for the drug recommender system has five phases, as shown in Figure 3: data preprocessing, sentiment mining, deep-learning-based sentiment classification, meta-heuristic optimization, and explainable recommendation. The proposed system uses a combination of NLP techniques, rule-based aspect mining, and deep learning models to extract local as well as contextual sentiment information from patient reviews. Aspect-level opinion mining is integrated with optimization-based weight assignment in order to incorporate the relative importance of aspects observed in patient-generated review data. Furthermore, a reliability-based ranking and explain ability module is incorporated to generate explainable and transparent patient-opinion-informed ranking results.

3.8.1 Drug Review Dataset

The patient review dataset from "Drugs.com" used in this study contains patient reviews of prescribed drugs based on real medical experiences shared on a health-related website. Each entry in this dataset contains the name of a drug, a related medical condition, a review text describing the patient's experience, and a corresponding rating value ranging from 1 to 10. The dataset contains reviews covering various medical conditions and drugs, thereby capturing multiple patient opinions about drug efficacy, side effects, and user satisfaction.

Table 4. Evaluation protocol and measures employed in this study

Component	Evaluation Method
Sentiment Classification	Accuracy, Precision, Recall, F1-score
Class-wise Evaluation	Confusion Matrix
Agreement Measurement	Cohen's Kappa Coefficient
Internal Baseline Comparison	CNN, Bi-LSTM, Capsule Network, MobileNet-CNN, Soft Voting Ensemble
State-of-the-Art Comparison	Recent Drug Review Sentiment Analysis Models
Recommendation Analysis	Top-K Drug Ranking Outputs
Optimization Analysis	WOA-Based Aspect Weight Evaluation

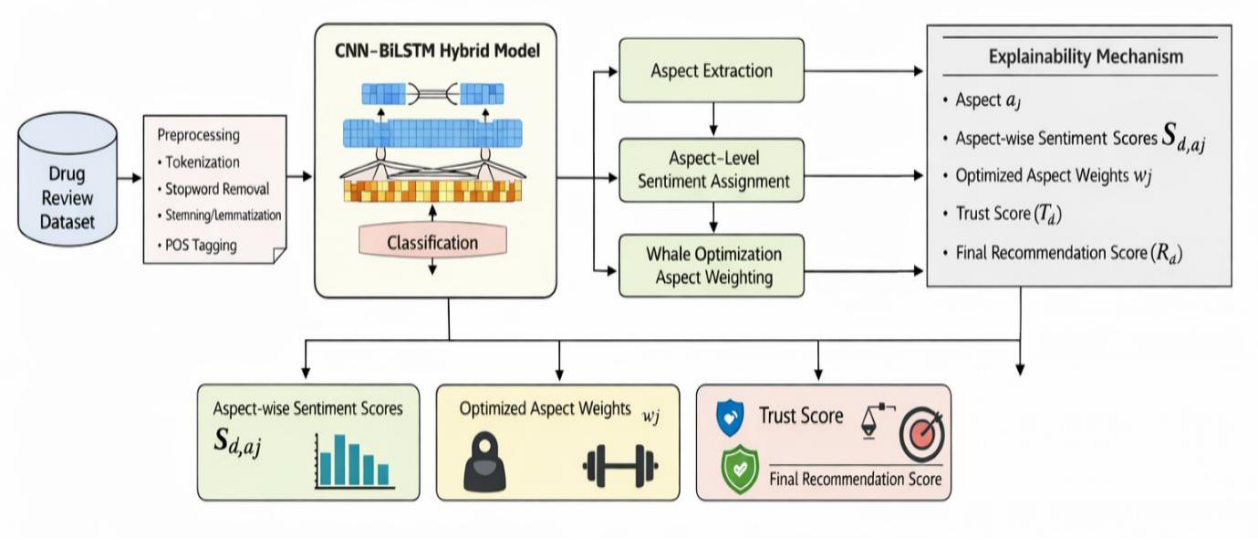

Figure 3. Architecture of the proposed framework incorporating a 100-dimensional embedding layer, 64 convolutional filters (kernel size = 5), and a BiLSTM layer with 64 hidden units.

Table 5. Dataset Statistics

Parameter	Value
Total Reviews	215,063
Training Reviews	161,297
Test Reviews	53,766
Rating Range	1–10
Positive Reviews	142,306
Neutral Reviews	19,185
Negative Reviews	53,572
Missing Condition Records	1,194

As shown in Table 5, the dataset is large, diverse, and real-world in nature, making it suitable for sentiment analysis, aspect-level opinion mining, and data-driven drug recommendation systems.

3.8.2 Data Preprocessing and Text Normalization

The initial stage of the framework focuses on pre-processing patient drug reviews to minimize noise and improve model performance. Drug review text may contain spelling errors, punctuation, other irrelevant symbols, and sentences of varying lengths. Pre-processing is therefore necessary to prevent poor model performance. The following are the steps in the pre-processing phase:

3.8.2.1 Text Cleaning

All reviews are converted to lowercase, as case variation may affect the output. Special characters, HTML tags, URLs, and non-alphabetic symbols are removed because they do not carry significant information for sentiment analysis. Extra spaces are removed to ensure a clean token sequence.

3.8.2.2 Tokenization

Each review is tokenized into words using a tokenizer that assigns an integer index to each unique word. This tokenization function converts text data into numerical values so that it can be processed by the neural network.

3.8.2.3 Padding and Sequence Length Standardization

As neural networks accept input sequences of fixed length, all tokenized reviews are either padded or reduced to a fixed maximum sequence length. When a review is shorter than this length, it is padded with zeros. When a review is longer than this length, it is truncated to retain the most significant information. This ensures that the text data is well formatted and normalized so that the subsequent deep learning models can perform well.

3.8.3 CNN-BiLSTM Model

This stage is responsible for automatically identifying the sentiment polarity expressed in patient

drug reviews. The convolutional and BiLSTM layers are jointly trained to learn sentence representations for extracting complex features, thereby improving sentiment classification performance. The sentiment classification model is a key component of the system because it provides the predicted sentiment labels for aspect-level sentiment analysis, optimization, and explainable drug recommendation. The hyper parameters of the CNN-BiLSTM model are given in Table 6.

Table 6. CNN-BiLSTM Hyperparameter Configuration

Parameter	Value
Tokenizer Vocabulary Size (num_words)	10,000
Embedding Dimension	100
Maximum Sequence Length	100
Convolution Filters	64
Kernel Size	5
Pooling Size	4
BiLSTM Hidden Units	64
LSTM Dropout	0.10
Recurrent Dropout	0.30
Optimizer	AdamW
Learning Rate	0.001
Weight Decay	1×10^{-5}
Batch Size	128
Training Epochs	10
Loss Function	Categorical Cross-Entropy
Output Activation	Softmax

3.8.4 Whale Optimization-Based Aspect Weighting

To perform sentiment analysis at the aspect level, an aspect extraction module is designed to extract the clinical aspects mentioned in the reviews. The CNN-BiLSTM classifier is used to predict the sentiment polarity of each extracted aspect instance. WOA is employed to learn optimal aspect weights.

Table 7. Whale Optimization Algorithm Configuration

Parameter	Value
Optimization Algorithm	Whale Optimization Algorithm (WOA)
Search Space	Continuous
Search Bounds	[0, 1]
Population Size	10
Maximum Iterations	30
Convergence Criterion	Maximum number of iterations
Optimization Objective	Optimize aspect weights for recommendation score computation

The WOA optimizes aspect weights using 10 candidate solutions and iterates up to a maximum of 30 times in the defined search space of [0, 1], as shown in Table 7. Once optimized, the aspect weights are combined with the corresponding trust scores to generate an overall recommendation score for each candidate drug.

3.8.5 Trust Score Computation and Drug Recommendation

A trust score is computed by integrating sentiment strength, review volume, and consistency. The drugs are then ranked as per Eq. (14), thereby generating the top-k recommendation list. The trust score calculated in this study was obtained by estimating the trustworthiness of each drug using the previously established trust estimation strategy. The final recommendation scores represent an aggregation of the aspect weights and trust scores computed during WOA. No additional trainable coefficients were introduced beyond those already built into the WOA optimization process.

3.8.6 Explainability Mechanism

An explainability layer is integrated to improve transparency and user trust. The aspect-based sentiment analysis for the drug recommender system is given in Table 8. For each recommended drug, the system reports:

- Aspect-wise sentiment scores S_{d,a_j}
- Optimized aspect weights w_j
- Trust score T_d
- Final recommendation score R_d

4. Experimental Results

The experimental results obtained from training the proposed framework are given in Figures 4 (a, b). Model evaluation is based on classification metrics such as accuracy, precision, recall, F1-score, the confusion matrix, and Cohen's Kappa coefficient.

Figure 4a shows that both the training and validation loss decrease monotonically during training, and the two curves remain very close to each other. This implies that the model is not overfitted. In Figure 4b, both the training and validation accuracy increase monotonically across epochs. In addition, the validation accuracy is slightly higher than the training accuracy at the end of training. This can be attributed to the regularization method applied in the model.

Based on the convergence behavior seen in Figures 4a and b, it can be concluded that the CNN-BiLSTM architecture can learn discriminative sentiment representations from the drug-review dataset.

Table 8. Aspect-based sentiment analysis for the drug recommender system using the Acetaminophen / butalbital / caffeine example

Component	Symbol	Description	Example Value
Aspect	a_j	Drug-related aspects considered in recommendation	Effectiveness, Side Effects, Safety, Dosage, Cost
Aspect-wise Sentiment Score	S_{d,a_j}	Sentiment polarity score of drug (d) with respect to each aspect (a_j)	Effectiveness = 2.000, Side Effects = 1.500, Safety = 1.800, Dosage = 1.700, Cost = 2.100
Optimized Aspect Weight	w_j	Weight assigned to each aspect after WOA optimization	Effectiveness = 0.105, Side Effects = 0.200, Safety = 0.180, Dosage = 0.190, Cost = 0.325
Weighted Aspect Contribution	$w_j \times S_{d,a_j}$	Contribution of each aspect to the final recommendation score	Effectiveness = 0.210, Side Effects = 0.300, Safety = 0.324, Dosage = 0.323, Cost = 0.6825
Trust Score	T_d	Reliability score computed from sentiment consistency and review confidence	8.251
Final Recommendation Score	R_d	Aggregated recommendation score obtained by combining the trust score with the weighted contributions of all aspects for ranking	15.17

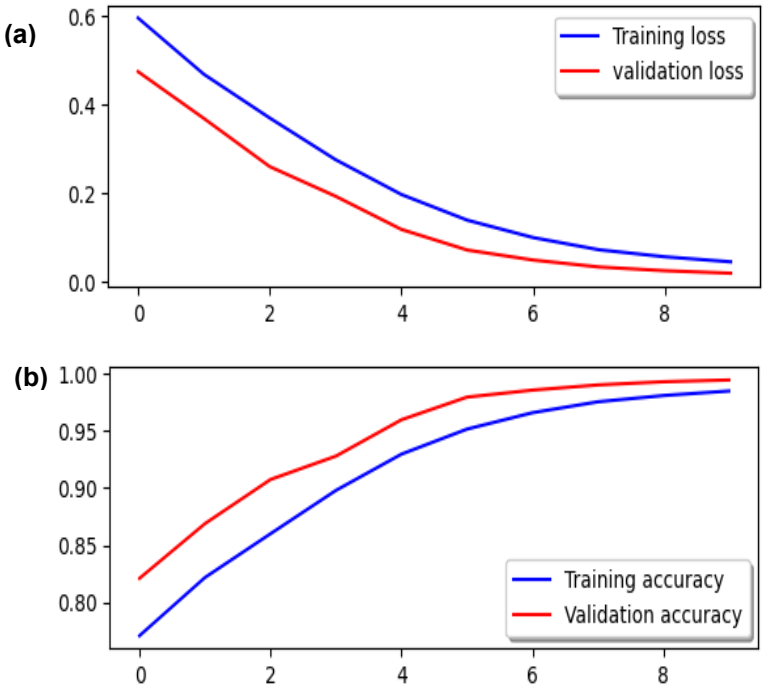

Figure 4 (a). Training and validation loss curves across epochs, **(b)** Training and validation accuracy curves across epochs.

The ability of the loss to continue decreasing steadily and of accuracy to continue increasing steadily indicates that the model is able to extract both local semantic patterns and long-range contextual dependencies. The small separation between the training and validation curves also indicates strong generalization properties and supports the usefulness of the dropout-based regularization procedure in mitigating overfitting while maintaining predictive performance.

The proposed classification model is further evaluated by the confusion matrix shown in Figure 5. It is evident from the matrix that most samples in each of the three classes—negative, neutral, and positive—are correctly classified and that only a few samples are misclassified, which implies that the classifier has learned the classes well and remains balanced. Specifically, samples from the neutral class are classified very accurately, although this is a notoriously difficult sentiment class because its language often contains highly subjective and ambiguous content along with implicit sentiment polarity.

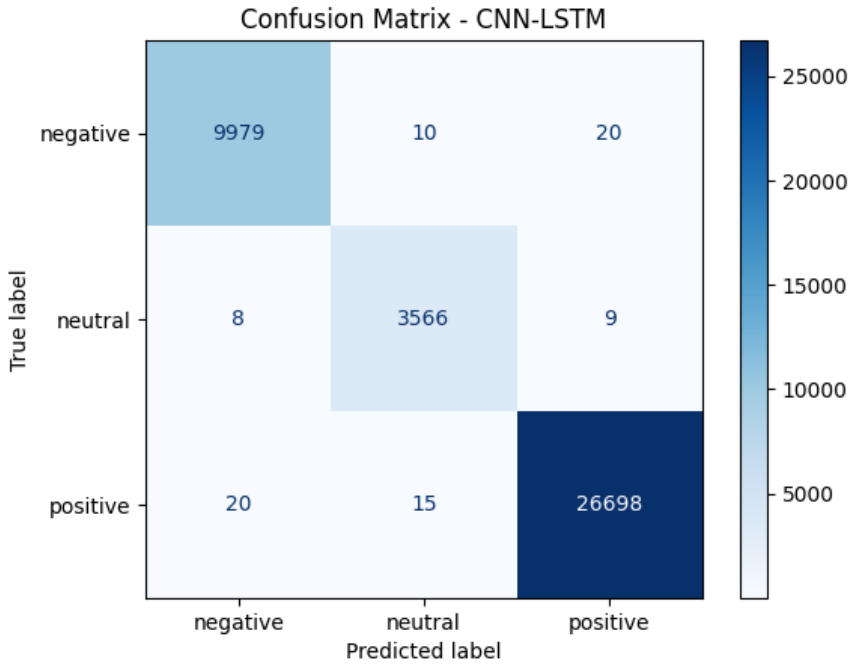

Figure 5. Confusion matrix showing sentiment classification.

Therefore, these results also demonstrate the ability of the proposed classifier to learn latent contextual relationships within the review text. Further analysis of the confusion matrix shows that the highest classification rate is achieved for the positive sentiment class because positive reviews generally contain clear expressions of treatment effectiveness and user satisfaction. The negative class also achieves a high classification rate due to the presence of clear descriptions of adverse effects and indicators of user dissatisfaction. The accuracy of the neutral sentiment class is also exceptionally high, although neutral reviews have traditionally been challenging to classify because they often include a nearly equal distribution of positive and negative expressions. The combination of local phrase-level feature extraction provided by the CNN layer and contextual dependency modeling provided by the Bi-LSTM layer allows the model to distinguish sentiments more clearly than traditional methods. The small number of misclassified documents between adjacent sentiment classes provides additional support for the reliability of the classifier proposed in this study.

Table 9. Precision, recall, and F1-score achieved by the proposed model for each sentiment class.

Class	Precision	Recall	F1-score
Negative	0.9972	0.9970	0.9971
Neutral	0.9930	0.9953	0.9941
Positive	0.9989	0.9987	0.9988
Accuracy			0.9980
Macro Avg	0.9964	0.9970	0.9967
Weighted Avg	0.9980	0.9980	0.9980

The high macro-averaged scores confirm that the model maintains uniform performance across all sentiment categories without bias toward majority classes. Based on the high average values shown in Table 9, the model provides a fairly even level of performance across each sentiment category and does not show strong bias toward the majority class. The high precision, recall, and F1-score for all three sentiment classes (negative, neutral, and positive) indicate very consistent learning across classes. The F1-score for the neutral class (i.e., 0.9941) is particularly noteworthy because the neutral class is challenging due to the ambiguity of neutral sentiment in healthcare reviews. Achieving this result demonstrates the ability of the CNN-BiLSTM architecture to recognize contextual information and resolve sentiment ambiguity. The close match between the weighted-average metrics and the overall accuracy value also indicates that model performance has not been adversely affected by class imbalance.

4.1 Comparative Analysis with Internal Baseline Models

To further validate the effectiveness of the proposed framework, its performance is compared with several internally implemented baseline models, including conventional deep learning architectures and ensemble approaches, as given in Table 10.

The proposed model obtains the highest accuracy, outperforming all baseline models by a clear margin. While MobileNet-CNN and Soft Voting ensembles show competitive performance, they remain inferior to the proposed architecture.

Table 10. Accuracy comparison with baseline models

Model	Accuracy
CNN	0.87
Bi-LSTM	0.85
Capsule Network	0.88
MobileNet-CNN	0.97
Soft Voting Ensemble	0.93
Proposed Model	0.998

Table 11. Comparative analysis of the proposed model based on macro-averaged precision, recall, F1-score, and accuracy.

Model	Precision (Macro Avg)	Recall (Macro Avg)	F1-score (Macro Avg)	Accuracy
CNN	0.79	0.77	0.78	0.87
Bi-LSTM	0.73	0.70	0.71	0.85
Capsule Network	0.87	0.86	0.86	0.88
MobileNet-CNN	0.97	0.95	0.96	0.97
Soft Voting Ensemble	0.93	0.92	0.92	0.93
Proposed Model	0.996	0.997	0.997	0.998

The performance gap can be explained by the fact that the two components of the proposed model leverage complementary strengths: CNNs extract local sentiment-bearing features, while BiLSTMs capture sequential contextual relationships among words in the review text. Therefore, separate models using only CNNs or only BiLSTMs cannot exploit both types of features simultaneously. Similarly, ensemble methods improve prediction stability, but they typically do not provide the same deep contextual understanding as the proposed model. This explains why the proposed model consistently achieves better classification performance across all evaluation metrics.

The macro-averaged comparison of precision, recall, F1-score, and accuracy is reported in Table 11 to ensure a fair evaluation across all sentiment classes.

Table 11 shows that the proposed model not only improves overall accuracy but also significantly increases macro-averaged precision, recall, and F1-score, confirming its robustness and balanced classification behaviour.

The performance improvements of the proposed framework are also shown in Table 11. Compared with CNNs, BiLSTMs, Capsule Networks, and Soft Voting Ensemble methods, the proposed architecture has substantially higher average precision, recall, and F1 metrics. The average improvement across all measures is important because it shows how well the model performs on a class-by-class basis rather than merely dominating majority classes. These results strongly indicate that combining CNNs, which provide local feature extraction, with Bi-LSTMs, which provide contextual modelling, yields richer representations of drug-review sentiments.

In addition to sentiment classification, the framework performs aspect-level sentiment analysis, estimates trust scores, and applies WOA-based aspect weighting in the recommendation module. These components do not directly affect sentiment classification accuracy because they operate after the classification step. However, they improve the recommendation process by enabling more precise aspect assessment, trust-sensitive scoring to lessen the impact of unreliable reviews, and WOA-based optimization of aspect importance. Thus, the entire framework is more robust in generating interpretable and patient-specific drug recommendations without sacrificing sentiment classification performance.

Unlike conventional recommendation datasets, the Drug Review Dataset adopted in this research does not include static opinion labels or expert-validated recommendation lists for each illness. Therefore, standard recommendation evaluation metrics such as Precision@K, Recall@K, or NDCG cannot be reliably computed. Accordingly, the recommendation component is assessed through condition-ranking experiments, trust-aware recommendation scores, aspect-wise contribution analysis, and reproducibility analysis across multiple executions. These experiments

show that the proposed framework produces patient-opinion-informed drug rankings consistently and interpretably while ensuring stable recommendation behavior.

4.2 Comparison with other State-of-the-Art Models used for Sentiment Analysis of Drug Reviews

The proposed model is compared with recent sentiment analysis approaches specifically applied to drug-review datasets. The comparative performance of the proposed method against recent drug-review sentiment analysis models, based on classification accuracy, is given in Table 12.

As shown in Table 12, deep learning models have outperformed classical machine learning models in the sentiment analysis of drug reviews. The classical Linear SVC with TF-IDF showed competitive performance with 93% accuracy, which may be attributed to its dependence on hand-engineered features, although such features lack contextual understanding. Transformer models such as BERT performed better, with an accuracy of 96%, because they use context-aware embeddings. However, CNN-BiLSTM models are able to learn n-gram features from the text while also modeling sequential dependencies, and they have achieved an accuracy of 97%. Other ensemble models and transformer-RNN hybrid models, such as RoBERTa + BiLSTM, have been reported to achieve an accuracy of 94.78%. The proposed CNN-BiLSTM model with metaheuristic optimization outperformed these state-of-the-art methods with 99.80% accuracy.

4.3 Recommendation Performance Evaluation

To assess the recommendation ability of the proposed framework, recommendation experiments are performed using representative medical conditions extracted from the drug-review dataset. For each condition, top-3 ranked drug recommendations are produced by integrating the calculated aspect-level sentiment scores, the trust score for the website from which patients shared their comments, and WOA-optimized aspect weights to formulate semantic recommendation reports. The example rankings demonstrate the ability of the proposed framework to combine multiple dimensions of patient reviews into transparent recommendation scores.

Table 12. Comparison with recent drug-review sentiment analysis models

Reference	Model / Approach	Technique Category	Reported Accuracy (%)
Al-Hadhrami *et al.* [9]	CNN+Bi-LSTM	Hybrid deep learning	97
Al-Hadhrami *et al.* [9]	CNN+Bi-LSTM (with GloVe variants)	Hybrid deep learning	87
Hayath Ali *et al.,* 2025 [32]	Gemini API + ML	Deep learning + NLP	93
Chaudhary *et al.* (2025) [37]	BERT	Transformer-based	96
Garg *et al.* [21]	Linear SVC + TF-IDF	Classical ML	93
Kim *et al.*,2025 [10]	RoBERTa + BiLSTM	Transformer hybrid	94.78
Patil *et al.* [38]	Domain BiLSTM	Hybrid deep learning	90.46
Proposed Model	CNN–BiLSTM + Meta-heuristic Optimization	Hybrid + Optimized deep learning	99.80

Table 13. Top-3 Recommended Drugs for the Headache Condition

Rank	Drug	Trust Score (T_d)	Final Recommendation Score (R_d)
1	Acetaminophen / butalbital / caffeine	8.251	15.17
2	Acetaminophen / dichloralphenazone / isometheptene mucate	6.824	13.233
3	Sumatriptan	6.655	12.481

Table 14. Top-3 Recommended Drugs for the Vomiting Condition

Rank	Drug	Trust Score (T_d)	Final Recommendation Score (R_d)
1	Zofran	7.853	14.226
2	Promethazine	6.808	12.629
3	Ondansetron	6.888	11.437

Table 15. Aspect-wise Contribution for the Top-3 Recommended Drugs

Drug	Effectiveness	Side Effects	Safety	Dosage	Cost
Acetaminophen/Butalbital/Caffeine	1.886	1.793	—	1.780	2.000
Acetaminophen/Dichloralphenazone/Isometheptene Mucate	1.894	1.788	2.000	2.000	2.000
Sumatriptan	1.821	1.776	—	1.816	2.000

Note: The symbol "—" indicates that no safety-related aspect was identified in the available patient reviews for the corresponding drug. Therefore, no safety sentiment score was assigned for that aspect during recommendation score computation.

Table 16. Component-wise Analysis of the Recommendation Framework

Framework Component	Contribution to Recommendation
CNN–BiLSTM	Classifies patient review sentiment
+ Aspect-Level Sentiment Analysis	Identifies sentiment for individual drug aspects (effectiveness, safety, side effects, dosage, and cost)
+ Trust Score Estimation	Reduces the influence of unreliable or inconsistent reviews during ranking
+ WOA-Based Aspect Weighting	Optimizes the relative importance of aspects to improve ranking consistency
Proposed Framework	Produces interpretable, patient-opinion-informed drug recommendations

The best drugs recommended by the proposed framework for the headache and vomiting conditions are shown in Tables 13 and 14. Ranking is determined by the final recommendation score, which is computed by aggregating aspect-level sentiment scores using trust-aware weighting optimized by WOA. The findings indicate that drugs with stronger effectiveness-related sentiment and higher trust scores tend to receive higher recommendation scores. After extracting patient opinions from online reviews, the recommendation component prioritizes drugs, as illustrated in these examples. Although Promethazine and Ondansetron have comparable trust scores, Promethazine achieves a higher final recommendation score because the final ranking is determined by a combination of the trust score and the weighted aspect scores optimized using WOA. The trust score adjusts the reliability of patient reviews, whereas the final recommendation score additionally reflects drug effectiveness, side effects, safety, dosage, and cost. Consequently, a drug with a slightly lower trust score may still obtain a higher final recommendation score if its overall weighted aspect performance is superior.

The aspects involved in the recommendation process for the top-ranked drugs are reported in Table 15. The scores for effectiveness, dosage, safety, and cost-related sentiment are combined with confidence weighting based on the WOA-derived overall aspect score in order to arrive at a higher-level recommendation after incorporating consistency-based loading. This improves interpretability by showing how each individual component contributes to the final recommendation.

To analyse the reproducibility of the recommendation framework, the recommendation procedure is repeated using both the trained CNN-BiLSTM model and the optimized WOA weights. Additionally, repeated executions of the algorithm generated identical top-3 recommendation lists, thereby indicating that the framework is able to produce stable and reproducible rankings under the same initial conditions.

The incremental contribution of each building block of the proposed recommendation framework is outlined in Table 16. The CNN-BiLSTM-based model classifies the sentiment of patient reviews into sentiment

polarities. The aspect-level sentiment analysis component builds on this by identifying sentiments associated with individual drug attributes, such as effectiveness, side effects, and dosage. The trust score component enhances recommendation reliability by reducing the impact of unreliable or inconsistent drug reviews, while WOA optimizes the relative weights of the various aspects used in ranking the drugs. As these components are applied after sentiment classification, they enhance the recommendation process without affecting the accuracy of sentiment classification itself.

5. Discussion

5.1 Analysis of CNN and BiLSTM

Our experimental results show that the proposed model, which combines CNN and BiLSTM, performs very well in sentiment classification of drug reviews. The CNN-BiLSTM-based model outperforms traditional deep learning models, ensemble models, and other state-of-the-art approaches across all evaluation measures. This high level of performance results from the ability of CNN and BiLSTM to capture different types of information. The CNN is effective at capturing local sentiment-laden phrases and features related to particular aspects, while the BiLSTM learns the long-term context in which a patient-generated healthcare review is written. Combining these two abilities—local feature extraction and sequential contextual learning—gives the CNN-BiLSTM-based model richer semantic representations than either CNN or any RNN alone.

5.2 Analysis Based on the Confusion Matrix

Another clear indicator of classification robustness is the confusion matrix, which shows consistently accurate classification of the negative, neutral, and positive sentiment classes. In particular, the neutral sentiment class achieves high precision and recall. Neutral sentiment is one of the most difficult classes to classify correctly in healthcare reviews because neutral reviews often contain mixed treatment experiences. In addition, neutral reviews often do not express strong positive or negative opinions regarding a treatment, usually contain dosage changes, and may include both treatment and control statements; therefore, they can contain many different types of modifying expressions. This demonstrates that the CNN-BiLSTM-based model is able to learn and accurately identify the subtle contextual and semantic properties of words in neutral reviews, thereby distinguishing them from polarized sentiments. While the model described here performs very well overall, most misclassified examples occur between adjacent sentiment classes, especially neutral-positive and neutral-negative reviews. One of the major contributors to these error cases is that many mixed or ambiguous reviews—for example, reviews containing conflicting experiences, confusion between effectiveness and side effects, or uncertainty about why an experience should be classified as positive or negative—reflect the inherently subjective and linguistically ambiguous nature of patient-generated healthcare reviews.

These results are consistent with prior research showing that hybrid CNN-BiLSTM systems can outperform standalone CNN and recurrent models because they capture both local semantic patterns in review wording through the CNN component and long-range contextual dependencies through the BiLSTM component. The findings therefore support the suitability of this hybrid deep learning approach for analyzing complex patient-generated drug review datasets. The proposed framework demonstrates the potential of analyzing large collections of patient-written medication reviews to understand public opinion about medications in a structured and explainable manner. By integrating aspect-level sentiment analysis, trust scores, and Whale Optimization Algorithm-based aspect weighting into the ranking process, the framework generates comparative medication rankings derived from patient perspectives. Such rankings may be useful for health services researchers and analysts as an opinion-aware supplementary resource for comparing medications.

At the same time, several limitations should be acknowledged. The study relies entirely on patient-generated drug reviews, which may contain subjective opinions, reporting bias, duplicate entries, missing information, and ambiguous descriptions. Review distributions may also be skewed because patients are more likely to submit feedback when they have very positive or very negative experiences, whereas moderate opinions are often underrepresented. In addition, the recommendation rankings were derived only from patient reviews and sentiment analysis without validation against clinical evidence, physician judgment, or verified treatment outcomes. Furthermore, the rule-based aspect extraction method depends on predefined dictionaries and extracted sentences, which may not adequately capture implicit medical concepts, domain-specific terminology, or previously unseen expressions.

6. Conclusion and Future Direction

This study presented an explainable drug recommendation framework that integrates aspect-based sentiment analysis, a hybrid CNN-BiLSTM classifier, trust-aware ranking, and Whale Optimization Algorithm-based aspect weighting to analyze large-scale patient drug reviews. The proposed model effectively captured both local semantic cues and long-range contextual dependencies in review text, enabling highly accurate sentiment classification and more informative aspect-level interpretation. By combining sentiment strength with optimized aspect importance and review consistency, the framework produced transparent opinion-aware rankings of medications for specific

conditions. The findings demonstrate that patient-generated review data can be transformed into a meaningful supplementary decision-support resource when supported by robust deep learning and optimization techniques. In particular, the framework moves beyond coarse polarity prediction by showing how effectiveness, side effects, dosage, safety, and cost can jointly shape recommendation outcomes in an interpretable manner. At the same time, the study recognizes that such rankings reflect patient opinion rather than clinically validated treatment guidance. Therefore, the proposed system should be considered a complementary analytical tool for researchers and healthcare analysts. Future work should strengthen clinical relevance through expert-annotated datasets, transformer-based aspect extraction, external evidence integration, and validation against real-world treatment outcomes. These advances would improve reliability, interpretability, and real-world usefulness in healthcare settings substantially.

References

[1] A.I. Stoumpos, F. Kitsios, M.A. Talias, Digital Transformation in Healthcare: Technology Acceptance and Its Applications. International Journal Environment Research and Public Health, 20(4), (2023) 3407. https://doi.org/10.3390/ijerph20043407

[2] B. Zhou, G. Yang, Z. Shi, S. Ma, Natural language processing for smart healthcare. IEEE Reviews in Biomedical Engineering, 17, (2025) 4-18. https://doi.org/10.1109/RBME.2022.3210270

[3] S. Eguia, C.L. Sánchez-Bocanegra, F. Vinciarelli, F. Alvarez-Lopez, Clinical decision support and natural language processing in medicine: A systematic literature review. Journal of Medical Internet Research, 26, (2024) e55315. https://doi.org/10.2196/55315

[4] S. ten Hoope, K. Welvaars, K. van Geijtenbeek, M. Klok-Everaars, S. van Schaik, F. Karapinar-Çarkit, Applying text-mining to clinical notes: the identification of patient characteristics from electronic health records (EHRs). BMC Medical Informatics and Decision Making, 25(1), (2025) 302. https://doi.org/10.1186/s12911-025-03137-x

[5] P. Nitiéma, Artificial intelligence in medicine: Text mining of healthcare workers opinions. Journal of Medical Internet Research, 25, (2023), e41138. https://doi.org/10.2196/41138

[6] K. Wisaeng, B. Muangmeesri, Deep learning for email threat intelligence: Feature importance and model performance in phishing detection. International Journal of Innovative Research and Scientific Studies, 8(5), (2025) 1853–1865. https://doi.org/10.53894/ijirss.v8i5.9302

[7] I. Aggarwal, S. Joseph, N. Jaganathan, A. Patel, V. Kumar, M. Devarapalli, Sentiment analysis in healthcare: A comparison of VADER, BERT, and Flair NLP models on patient reviews of pain management physicians. Cureus, 17(7), (2025) e88902. https://doi.org/10.7759/cureus.88902

[8] H. Liu, H. Qiao, X. Shi, M. Shang, (2022) Aspect-aware Asymmetric Representation Learning Network for Review-based Recommendation. International Joint Conference on Neural Networks (IJCNN), IEEE, Italy. https://doi.org/10.1109/IJCNN55064.2022.9892559

[9] S. Al-Hadhrami, T. Vinko, T. Al-Hadhrami, F. Saeed, S. Qasem, Deep learning-based method for sentiment analysis for patients' drug reviews. PeerJ Computer Science, 10, (2024) e1976. https://doi.org/10.7717/peerj-cs.1976

[10] T.H. Kim, A. Aldrees, D.A. AlHammadi, M. Umer, T. Saidani, S. Alsubai, I. Ashraf, MediNet: ensemble transfer learning approach for classification of medical drugs-related text reviews using significant combined-embeddings. BioData Mining, 18(1), (2025) 71. https://doi.org/10.1186/s13040-025-00448-7

[11] S.V. Kumar, J.S. Chohan, K. Kalita, WhaleOptNB: A Method for Automated Biomedical Text Document Classification. Intelligent Computing and Optimization, (2025) 174–184. https://doi.org/10.1007/978-3-031-73324-6_18

[12] B. Srishailam, V.V. Lavanya, V. shivani, O. Ashok, R.S. Kiran, Integrating Sentiment Analysis and Machine Learning for Patient-Centric Drug Recommendation Systems. International Journal of Computer Engineering in Research Trends, 11(1s), (2024) 9–15. https://doi.org/10.22362/ijcert/2024/v11/i1/v11i1s02

[13] P. Sequeira, S. Begum, MI-Powered Emergency Drug Assistance for Critical Healthcare Needs. Journal of Emerging Technologies and Innovative Research (JETIR), 12(7), (2025) 77-83. https://www.jetir.org/papers/JETIRGX06015.pdf

[14] Y. Wang, A. Liu, J. Yang, L. Wang, N. Xiong, Y. Cheng, Q. Wu, Clinical knowledge-guided deep reinforcement learning for sepsis antibiotic dosing recommendations. Artificial intelligence in medicine, 150, (2024) 102811. https://doi.org/10.1016/j.artmed.2024.102811

[15] J. Wang, B. Xu, Y. Zu, (2021) Deep learning for Aspect-based Sentiment Analysis. International Conference on Machine Learning and Intelligent Systems Engineering (MLISE), IEEE, China. https://doi.org/10.1109/MLISE54096.2021.00056

[16] J. Lin, M.K. Najafabadi, (2023) Aspect Level Sentiment Analysis with CNN Bi-LSTM and Attention Mechanism. 26th ACIS International

Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter), IEEE, China. https://doi.org/10.1109/SNPD-Winter57765.2023.10466355

[17] F. Di Martino, F. Delmastro, Explainable AI for clinical and remote health applications: a survey on tabular and time series data Artificial Intelligence Review, 56(6), (2023) 5261–5315. https://doi.org/10.1007/s10462-022-10304-3

[18] Y. He, (2023) BERT-CNN-BiLSTM: A Hybrid Deep Learning Model for Accurate Sentiment Analysis. IEEE 5th International Conference on Power, Intelligent Computing and Systems (ICPICS), IEEE, China. https://doi.org/10.1109/ICPICS58376.2023.10235335

[19] Q. Li, X. Li, B. Lee, J. K. Kim, A Hybrid CNN-Based Review Helpfulness Filtering Model for Improving E-Commerce Recommendation Service. Applied Sciences, 11, (2021) 8613. https://doi.org/10.3390/app11188613

[20] H. Vanam, MedRecommenderX: a scalable sentiment-aware framework for personalized drug recommendations. Network Modeling Analysis in Health Informatics and Bioinformatics, 14(1), (2025) 165. https://doi.org/10.1007/s13721-025-00664-5

[21] S. Garg, (2021) Drug Recommendation System based on Sentiment Analysis of Drug Reviews using Machine Learning. 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), IEEE, India. https://doi.org/10.1109/Confluence51648.2021.9377188

[22] S.K. Nayak, M. Garanayak, S.K. Swain, S.K. Panda, D. Godavarthi, an Intelligent Disease Prediction and Drug Recommendation Prototype by Using Multiple Approaches of Machine Learning Algorithms. IEEE Access, 11, (2023) 99304 – 99318. https://doi.org/10.1109/ACCESS.2023.3314332

[23] K. Zhang, H. Qian, Q. Liu, Z. Zhang, J. Zhou, J. Ma, E. Chen, (2021) Sifn: A sentiment-aware interactive fusion network for review-based item recommendation. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management, 3627-3631. https://doi.org/10.1145/3459637.3482181

[24] B. Omarov, M. Baikuvekov, D. Sultan, N. Mukazhanov, M. Suleimenova, M. Zhekambayeva, Ensemble Approach Combining Deep Residual Networks and BiGRU with Attention Mechanism for Classification of Heart Arrhythmias. Computers, Materials and Continua, 80(1), (2024) 341–359. https://doi.org/10.32604/cmc.2024.052437

[25] Y. Ma, Q. Li, A weakly-supervised extractive framework for sentiment-preserving document summarization. World Wide Web, 22(4), (2019) 1401–1425. https://doi.org/10.1007/s11280-018-0591-0

[26] M. Kumar, T. Ali, J. Anguera, Suman Lata Tripathi (2025) Emerging Technologies in AI, Computation, Communication, and Cybersecurity. CRC Press, London. https://doi.org/10.1201/9781003739791

[27] L. Fu, P. Liang, Z. Rasheed, Z. Li, A. Tahir, X. Han, Potential Technical Debt and Its Resolution in Code Reviews: An Exploratory Study of the OpenStack and Qt Communities. International Symposium on Empirical Software Engineering and Measurement, IEEE Computer Society, (2022) 216 – 226. https://doi.org/10.1145/3544902.3546253

[28] P. Durga, D. Godavarthi, S. Kant, S.S. Basa, Aspect-based Drug Review Classification Through a Hybrid Model with Ant Colony Optimization Using Deep Learning. Discover Computing, 27(1), (2024). https://doi.org/10.1007/s10791-024-09441-w

[29] E. Hasan, C. Ding, (2023) Aspect-Aware Multi-Criteria Recommendation Model with Aspect Representation Learning. IEEE International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), IEEE, Italy. https://doi.org/10.1109/WI-IAT59888.2023.00043

[30] Z. Li, W. Cheng, R. Kshetramade, J. Houser, H. Chen, W. Wang, (2021) Recommend for a Reason: Unlocking the Power of Unsupervised Aspect-Sentiment Co-Extraction. Findings of the Association for Computational Linguistics: EMNLP, 763–778. https://doi.org/10.18653/v1/2021.findings-emnlp.66

[31] V.S.K. Koppisetti, The Role of Explainable AI in Building Trustworthy Machine Learning Systems. ESP International Journal of Advancements in Science & Technology (ESP-IJAST), 2(2), (2024) 16-21.

[32] S.D.N. Hayath Ali, G.D. Kumar, S. Lakshmi, D.C. Subhashree, M.M. Harshitha, Smart Drug Recommendation System Using Sentiment Analysis of Drug Reviews. International Journal for Research in Applied Science Engineering Technologu, 13(9), (2025) 864–867. https://doi.org/10.22214/ijraset.2025.74004

[33] R. Eden, A. Burton-Jones, J. Grant, R. Collins, A. Staib, C. Sullivan, Digitising an Australian University Hospital: Qualitative Analysis of Staff-Reported Impacts. Australian Health Review, 44(5), (2020) 677–689. https://doi.org/10.1071/AH18218

[34] T. Sengodan, S. Misra, M. Murugappan, (2025) Advances in Electrical and Computer Technologies, CRC Press, London. https://doi.org/10.1201/9781003515470

[35] R.F. Meem, K.T. Hasan, (2023) Improving Sentiment Analysis in Online Course Reviews with BERT and Transformer Attention Mechanism. Research Square. https://doi.org/10.21203/rs.3.rs-3741963/v1

[36] M.M. Rahman, A.I. Shiplu, Y. Watanobe, M.A. Alam, RoBERTa-BiLSTM: A Context-Aware Hybrid Model for Sentiment Analysis. IEEE Transactions on Emerging Topics in Computational Intelligence, 9(6), (2025) 3788–3805. https://doi.org/10.1109/TETCI.2025.3572150

[37] A. Chaudhary, S. Pokhrel, S. Ganesan, P.B. Shah, N. Somasiri, (2025) Drug Review Sentiment Analysis: Applying Transformer-Based Models for Enhanced Healthcare. Journal of Data Science and Intelligent Systems. https://doi.org/10.47852/bonviewJDSIS52024468

[38] A.K. Patil, M.V Nikum, Drugs Review Sentiment Analysis. International Journal of Advanced Research in Computer and Communication Engineering, 14(10), (2025) 238-245.

Authors Contribution Statement

Suruchi Chawla: Conceptualization, Methodology, Resources, Writing - Review & Editing. Aakanksha: Validation, Writing-Review Editing, Supervision. Madhu Rani: Methodology, Software, Formal Analysis, Investigation, Resources, Data curation, Writing-Original Draft, Visualization. All the authors have read and agreed to the published version of the manuscript.

Funding

The authors declare that no funds, grants or any other support were received during the preparation of this manuscript.

Declaration of AI Use

During the preparation of this manuscript, the authors used ChatGPT (OpenAI) solely to assist with language editing, including grammar, spelling, sentence structure, and readability. The tool was not used to generate scientific content, research data, analysis, interpretation and conclusions. All output was critically reviewed and edited by the authors, who take full responsibility for the final content of the manuscript.

Competing Interests

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Data Availability

The data supporting the findings of this study can be obtained from the corresponding author upon reasonable request.

Has this article screened for similarity?

Yes