ADROID Conference Series: Engineering and Technology
Volume 2, Issue 3, pp.245-257, 2026

Abstract-Based Sentiment Analysis with Drug Recommendation System

Bhupesh Kumar Singh, Shashwat Shukla, Aayush Jangid, Ravindhar Choudhary

Amity School of Engineering and Technology

Amity University Rajasthan Jaipur

bksingh@jpr.amity.edu, shashwat.shukla@s.amity.edu, aayush.jangid@s.amity.edu, ravindhar.choudhary@s.amity.ed u

Abstract. The rapid growth of the internet has enabled a geometric increase in user-generated healthcare content, with patients often writing reviews of their experiences with the drugs they are taking. Such reviews are extremely important for understanding how drugs work in the body. Thus, a framework is proposed to develop a system for recommending drugs based on medical condition classifications using Sentiment Analysis. The system is based on Natural Language Processing techniques such as data cleaning, tokenisation, and lemmatisation to convert unstructured reviews to machine-understandable structures. Various system architectures are proposed, including traditional machine learning-based structures and deep learning-based structures such as CNN, LSTM, and BERT. Utilising different datasets made available via various online sources, such as Drugs.com, experimental results establish the efficacy of the suggested framework, with an accuracy of 94.4% for condition classification using the Passive Aggressive Classifier, and 94.85% for sentiment prediction using Logistic Regression. The proposed engine filters out the highest-rated drugs based on patient-validated "useful counts" and sentiment analysis. This study suggests an effective tool for public health surveillance through sentiment analysis, aiding individuals in choosing from extensive medical treatments.

Keywords: Sentiment Analysis, Bidirectional Long Short-Term Memory (BiLSTM), Drug Recommendation System, Deep Learning, Natural Language Processing, Healthcare Text Mining.

1. INTRODUCTION

The recent surge in user-generated content on online healthcare platforms has witnessed an unprecedented volume of patient-generated drug reviews that describe their actual experience in using various drugs, which provides useful patient-related information on drug efficacy, side effects, or patient satisfaction. This area has traditionally been underreported or reported later in the drug life cycle in a timely manner through formal drug trials. [1] [2] [3] As a result, the domain of automated analysis of human sentiment and intelligent systems of recommendation of drugs has emerged as holding promise to facilitate clinical decision-making, to create awareness among the population, and to carry out post-marketing surveillance of drugs. [4] [5] [36]

Sentiment analysis, also known as opinion mining, is one of the basic but important tasks in the domain of Natural Language Processing, through which subjective information is identified and classified. [1] [7] Sentiment analysis faces unique challenges in the domain of health care due to informal language, domain-specific terminology, negations, and word polarities that are dependent on context. [8] [9] Early methods for drug review sentiment analysis relied on lexicon-based or traditional machine learning methods such as Naïve Bayes and SVMs, which were limited by handcrafted features and poor contextual understanding. [10] [11] However, with the advent of deep learning, recurrent neural networks, particularly Long Short-Term Memory (LSTM) and Bidirectional LSTM, have proven to achieve state-of-the-art results by successfully handling large contextual relationships in sequential text. [12] [13]

Recent studies have shown that BiLSTM and hybrid CNN–BiLSTM architectures outperform the conventional models in the tasks of sentiment classification of patient drug reviews with higher accuracy and robustness across diverse datasets. [2] [14]

[15] The bidirectional benefit of the BiLSTM enables the model to take advantage of past as well as future contextual elements, making the approach particularly appropriate for comprehending subtle sentiment as expressed in patient narratives. [2] [16][35] Based on these results, this study uses a sentiment analysis model that is based on the BiLSTM model to effectively classify the sentiments of the reviews given by patients.

In addition to this, another major challenge comes from the transformation of the obtained opinions into recommendations. Drug recommendation systems try to provide recommendations regarding prescription of medications based on patient conditions, preferences, and past responses. This helps to reduce information overload. [17] [18][39] Traditional Recommender Systems have typically made use of collaborative filtering-based methodologies, while others have leveraged content-based filtering. However, when it comes to the

medical domain, these methodologies have proven to be ineffective. [19] [20] Accordingly, hybrid drug recommendation models that incorporate semantic similarity, sentiment, and effectiveness have received growing attention. [21][32]

In this research, a new hybrid drug recommendation system is proposed, which comprises Term Frequency-Inverse Document Frequency (TF-IDF), Sentence-BERT (SBERT) embeddings, and SVM-based cosine similarity methods to incorporate both lexical and semantic similarity between patient queries and drug reviews. [22] [23] [24] To increase the reliability of recommendations, Bayesian smoothing is utilized over past drug success rates. [25][31] It reduces bias that might be introduced by few or unbalanced review distributions. Additionally, a top-k ranking technique is used in which consistently top-rated medicines with high sentiment scores in addition to past success rates are prioritized. [17] [21]

The prime research contributions lie in these three aspects:

1. The development of a BiLSTM-based sentiment analysis model suited to patient drug reviews
2. The design of a hybrid framework integrating TF-IDF, SBERT, SVM - cosine similarity, and Bayesian - smoothed historical success rates.
3. The integration of sentiment-aware top-k ranking to provide reliable drug recommendations.

By bridging the gap between deep learning methods of sentiment analysis and a multi-stage evidence-based recommendation strategy, this research can contribute to the development of healthcare-based NLP and recommender systems, assisting in intelligent and data-driven approaches for recommending drugs to patients, as depicted through their experiences.

2. LITERATURE REVIEW

Initially, most of the research done on sentiment analysis had been within the realm of general data formats like reviews on products and movies. In fact, the research done by Pang et al. [1] and Turney [10] proved the efficiency of machine learning and semantic orientation in the classification of data. Though the techniques seemed to be working well for general data sets, in the case of health care data, there had been a problem due to the informal nature and the context-dependent polarity of the data.

[8] [9][30]

Initial efforts towards drug review analysis used lexicon based and rule-based methods. Na et al. [11] presented a rule-based sentiment classification approach based on linguistic dependency structures and medical lexicons. While systems like these are effective in narrow, controlled environments, they were certainly not easily scalable and relied heavily on domain expertise. Traditional machine learning classifiers like Naïve Bayes, Logistic Regression and Support Vector Machines remained moderately successful but needed hand-designed features desperately. [26] [16][38]

However, with the advent of deep learning, new architectures began outperforming conventional architectures, using machines to automatically learn contextual representations. Long Short-Term Memory (LSTM) networks, introduced by Hochreiter and Schmidhuber in [12], solved the gradient vanishing problem and were used for modelling long-range dependencies. Bidirectional LSTM (BiLSTM) improved this capacity by using both past and future context. [13]

Several studies conclude that BiLSTM-based architectures outperform for drug review sentiment analysis. Colón-Ruiz and Segura-Bedmar [4] performed an extensive comparison among CNN, LSTM, BiLSTM, and BERT-based models and concluded that BiLSTM presents a good balance between accuracy and computational cost. Al-Hadhrami et al. [2] showed the high accuracy of patient drug review datasets using a hybrid BiLSTM–CNN model, further reinforcing the potency of bidirectional contextual learning. Survey studies by Zhu et al. [16][41] also pointed to BiLSTM as one of the dominant architectures for aspect-based sentiment analysis tasks within the healthcare domain.

These findings indeed provide the motivation for using a BiLSTM-based sentiment classifier in this work to capture subtle sentiment patterns robustly in both medical abstracts and patient-authored drug reviews.

Feature representation is an important aspect of not only sentiment classification but also of recommendation systems. Statistic approaches, such as Term Frequency-Inverse Document Frequency (TF-IDF), have remained prominent in representing the discriminative features, especially using linear models. [22] [27] However, the TF-IDF method does not take semantics into account, which is of primary importance for text understanding in the medical field.

Distributed word representation techniques, such as Word2Vec [22][37] and GloVe [28], have also been beneficial in advancing semantic representation capacity, along with contextual similarity. The transformer models have taken this capability to an entirely new level. BERT specifically has introduced deep context embeddings, which was a groundbreaking contribution towards achieving state-of-the-art in all NLP tasks. Sentence-BERT (SBERT) optimized BERT further for sentence-level semantic similarity, creating sentence embeddings for cosine similarity calculation. [24][42][36]

The latest healthcare NLP models use hybrid representations; these models use a combination of both sparse vector representations using the TF-IDF technique and dense vector representations for capturing specific elements and semantics, respectively. [16] [14] This approach to hybrid representations was considered while designing the present system, which combines entities such as TF-IDF, SBERT embeddings, and SVM cosine similarity for the purpose of semantic matching.

Recommender systems in healthcare try to support patients and clinicians by recommending proper medication, taking previous experiences and outcomes into consideration. Traditional collaborative filtering and content-based recommender systems show only limited applications because of data sparsity, safety concerns, and lack of interpretability in medical domains. [19] [20][34]

Recent studies advocate hybrid recommendation models that incorporate sentiment analysis, semantic similarity, and historical effectiveness. Rathnasekara and Wijenayake [6] proposed a sentiment-driven drug recommendation framework combining condition classification with patient feedback. Garg [21][35] demonstrated that machine learning–based drug recommenders benefit significantly from sentiment-aware ranking mechanisms.

Bayesian smoothing is used in recommendation systems to give stable estimates of success rates for rarely reviewed items, to address biases introduced by uneven review distributions. [25][32] The result obtained in top-k ranking further enhances usability by prioritizing the most reliable and relevant recommendations. [22] [29] These pieces of research together inspire the Bayesian-smoothed, similarity-based top-k ranking model proposed in this research.

Despite considerable success, the existing studies still deal with sentiment analysis and recommendations separately. Few systems integrate deep contextualized sentiment modelling with semantic similarity and probabilistic ranking mechanisms into one single system. Furthermore, many transformer-based solutions have high computational overheads, limiting practical deployments. [4][40][33]

To address these gaps, this work suggests using a computationally efficient architecture that is a combination of:

- Sentiment Analysis using BiLSTM for Context Understanding
- TF-IDF and SBERT embeddings for lexical and semantic similarity,
- SVM-based cosine similarity for discriminative matching, and
- Bayesian Smoothed Historical Success Rates with top-k Ranking for Robust Drug Recommendation.

This integrated approach will significantly contribute to progressing recommender systems in healthcare by combining sentiment intelligence with evidence-based personalization.

3. METHODOLOGIES

A. Abstract-Based Sentiment Analysis

The proposed design is a multistage pipeline, and these stages include data preparation, context summarization based on TF-IDF, sequence tokenization, and classification using a Deep Learning technique.

i. Data Acquisition and Data-Preprocessing

The text data used here were collected from drug reviews and satisfaction ratings. To make the classification easier, the satisfaction ratings $S \in [1, 5]$are categorized into a sentiment label

$$L = \begin{cases} \text{Negative}, & \text{if } S \leq 2 \\ \text{Neutral}, & \text{if } S = 3 \\ \text{Positive}, & \text{if } S \geq 4 \end{cases}$$

The initial cleaning involved removing HTML tags, URLs, and non-alphabetic characters using regular expressions.

ii. Hybrid TF-IDF Context Summarization

A TF-IDF summarization layer was implemented to minimize computing noise and concentrate on sentiment-containing keywords.

Term Frequency-Inverse Document Frequency (TF-IDF): The frequency of a word $w$in a document d within a set of documents $D$is defined as

$$\text{TF} - \text{IDF}(w, d, D) = tf(w, d) \times \log\left(\frac{N}{df(w, D)}\right)$$

where $tf(w, d)$ is the frequency of $w$in d, N is the total number of documents, and $df(w, D)$ is the number of documents with w.

The top 35% of tokens in each review, ranked by their TF-IDF scores, were retained to form a "summarized" review. This reduces the sequence length while preserving high-variance features.

iii. Text Vectorization and Embedding

The summarized text is passed through a tokenizer to convert the words into integer sequences. Let

$$x = [w_1, w_2, \dots, w_n]$$

be a sequence of word indices.

This is then passed to an Embedding Layer, which maps each word to a high-dimensional vector space:

$$E(w) \in \mathbb{R}^k$$

where k is the embedding dimension (i.e., 128).

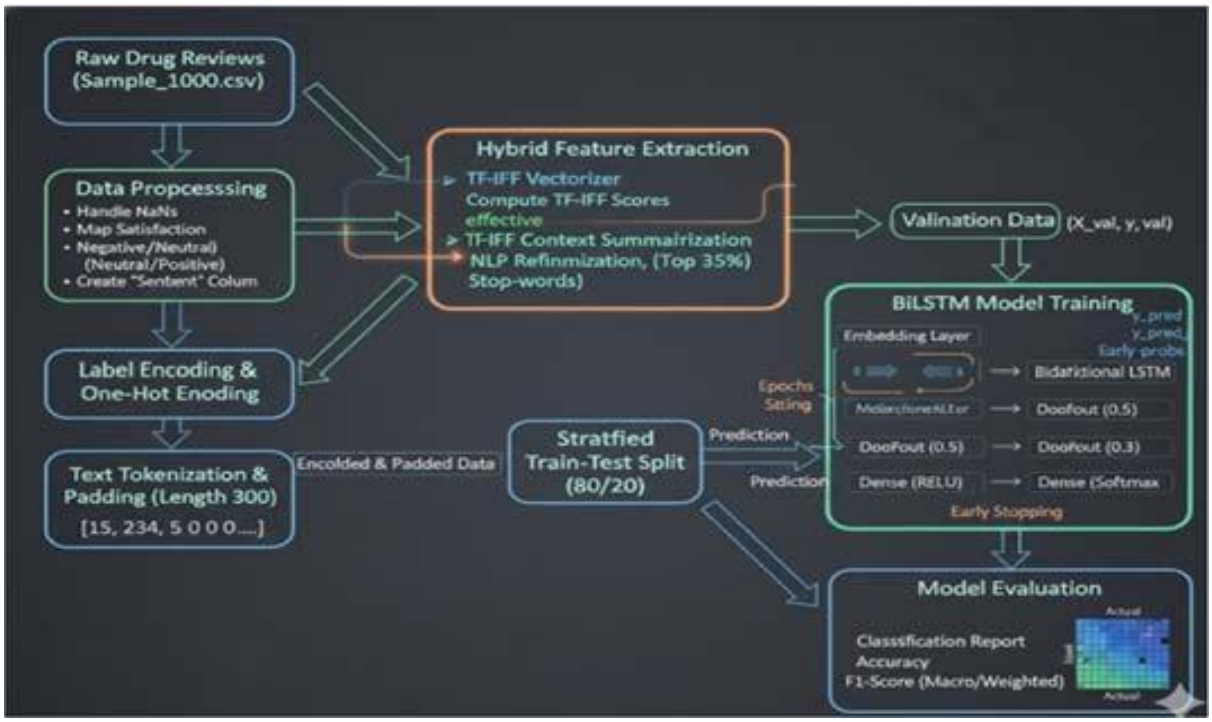

Fig. 1. Proposed Hybrid Recommendation Framework. The system utilizes dual-pipeline architecture: (1) a Satisfaction Classifier for real-time sentiment analysis of user text and a Prototype-Based Recommender that maps user conditions and severity to drug clusters using SBERT semantic embeddings and Bayesian-smoothed historical priors.

iv. BiLSTM Model Architecture

The core of our model is a bidirectional LSTM (BiLSTM). Unlike regular LSTMs, BiLSTMs process the input sequences from both the left $(\overrightarrow{h_t})$and right $\left(\overleftarrow{h_t}\right)$, which helps to consider the long-range dependencies within a word. The hidden state of the cell at position $t$is typically given by

$$H_t = \left[\overrightarrow{h_t} \oplus \overleftarrow{h_t}\right]$$

where $\overrightarrow{h_t}$and $\overleftarrow{h_t}$represent the forward and backward hidden states, respectively, and $\oplus$ denotes the concatenation operation. This combined representation captures contextual information from both preceding and succeeding words, enabling improved semantic understanding. The concatenated hidden representation is subsequently passed through a dropout layer to reduce overfitting, followed by a fully connected dense layer with ReLU activation and a SoftMax output layer to perform three-class sentiment classification.

where $\oplus$ denotes the concatenation. This is followed by a dropout layer to prevent overfitting, a dense layer with ReLU activation, and a final SoftMax layer for 3-class classification.

v. Optimization and Error Metrics

Because this is a multi-class classification problem, Categorical Cross-Entropy was used as the objective function to minimize. The error or loss Jcan be computed as

$$J(\theta) = -\sum_{i=1}^{M}\sum_{j=1}^{C} y_{i,j}\log(\hat{y}_{i,j})$$

Where:

- M is the number of samples.
- C is the number of classes (3).
- $y_{i,j}$ is a binary indicator (0 or 1) if class label $j$is the correct classification for observation i.
- $\hat{y}_{i,j}$ is the predicted probability of observation $i$belonging to class j.

vi. Error and Performance Metrics

To assess the robustness of the model against class imbalance, the following metrics were used:

Accuracy: This is the ratio of correctly predicted observations to total observations.

F1-Measure Macro: This is the arithmetic average of the respective F1-scores for each class. Thus, it is an unbiased measure that estimates the performance of minority classes, such as neutral.

$$F1\text{-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

vii. Optimizer

The model weights θ are updated using the Adam Optimizer, which uses adaptive learning rates for each parameter:

$$\theta_{t+1} = \theta_t - \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon}$$

where $\hat{m}_t$and $\hat{v}_t$are estimates of the first and second moments of the gradients, respectively.

B. Analytical Modelling

The proposed system uses a highly efficient binary sentiment classification mechanism for pharmaceutical product reviews. Its structure is based on a multiphase process involving data filtration, two-path feature detection, and linear classification.

1) Data Pre-Processing and Labelling

The raw data included satisfaction values Son a scale of 1–5. To ensure maximum confidence within the model, we used a Binary Classification technique to filter out 'neutral' data. The values of Ywere assigned as follows:

$$Y = \begin{cases} \text{Negative}, & \text{if } S \in \{1, 2\} \\ \text{Positive}, & \text{if } S \in \{4, 5\} \end{cases}$$

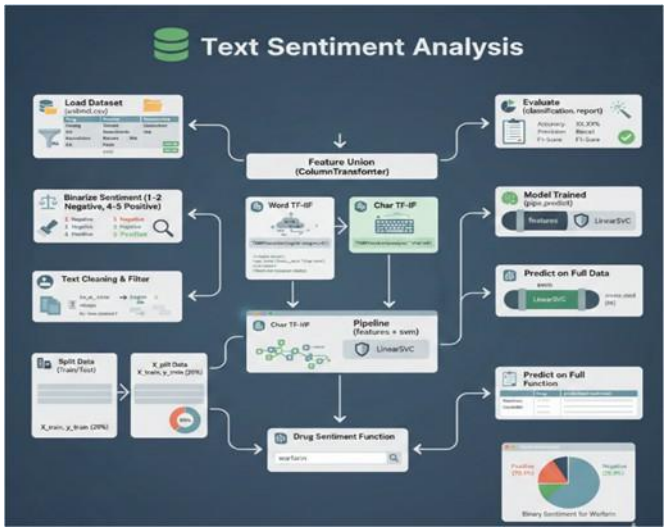

Fig. 2. Proposed Hybrid Recommendation Framework. The system utilizes dual-pipeline architecture: (1) a

Textual noise was reduced through the process of lowercase conversion, removal of HTML tags, and filtering reviews with lengths of less than 25 characters to ensure semantic density.

2) Feature Engineering: Dual TF-IDF Vectorization

We use the Feature Union technique to combine the word and character levels.
This method allows the model to store the semantic meaning of the words as well as the "morphological nuances such as prefixes/suffixes frequently embedded within words in medicine."
The importance of each feature was calculated using the **Term Frequency-Inverse Document Frequency (TF-IDF)** weight.
For term t in document d:

$$TF\ -\ IDF(t,d) = tf(t,d) \times \log\left(\frac{N}{df(t)}\right)$$

Where

- $tf(t,d)$ is the frequency of term $t$in document d.
- N is the total number of reviews in the corpus.
- $df(t)$ is the number of reviews having term t.

To further improve performance, we apply **Sublinear TF Scaling**, replacing $tf$with $1 + \log(tf)$to prevent highly frequent words from dominating the feature space.

3) Classification Model: Linear SVC

The classification was performed using a **Linear Support Vector Classifier (Linear SVC)**. The approach is to find the hyperplane that maximizes the class margin between the "positive" and "negative" classes.
The optimization problem is given by

$$\min_{w,b} \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{n} \xi_i$$

Subject to:

$$y_i(w^T x_i + b) \geq 1 - \xi_i,\ \xi_i \geq 0$$

Where:

- w is the weight vector and b is the bias.
- C is the regularization parameter (set to 5.0 in this study) which controls the trade-off between margin maximization and classification error.
- ξ_i are the slack variables being the error term for misclassified or margin-violating samples.

4) Evaluation and Error Metrics

To measure the predictive power and manage potential class imbalances, we employ three primary metrics derived from the Confusion Matrix.

i. Error Terms

The model performance was based on the counts of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN):

- **Accuracy:** The overall correctly predicted proportion:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precision and Recall:** Precision measures the exactness of the model, whereas Recall measures the completeness:

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

- **F1-Score:** The harmonic mean of Precision and Recall, providing a single metric for classification health:

$$F1\text{-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

ii. Class Weighting

Given the inherent imbalance in drug reviews, we implemented **Balanced Class Weights**. The weight for class $j$is calculated as follows:

$$W_j = \frac{N}{n_{\text{classes}} \times n_j}$$

where:

- N is the total number of training samples.
- n_{classes} is the total number of classes.
- n_j is the number of samples belonging to class j.

This adjustment ensures that the error term in the SVM objective function is penalized more heavily for the minority class, preventing the model from developing a majority-class bias.

C. Recommendation System

The proposed architecture includes a hybrid framework that aims to fulfill the tasks of patient satisfaction evaluation and pharmaceutical intervention recommendation. Considering this requirement, the method was grouped into three major phases: Feature Engineering, Sentiment Classification, and Bayesian Smooth Similarity Recommendation.

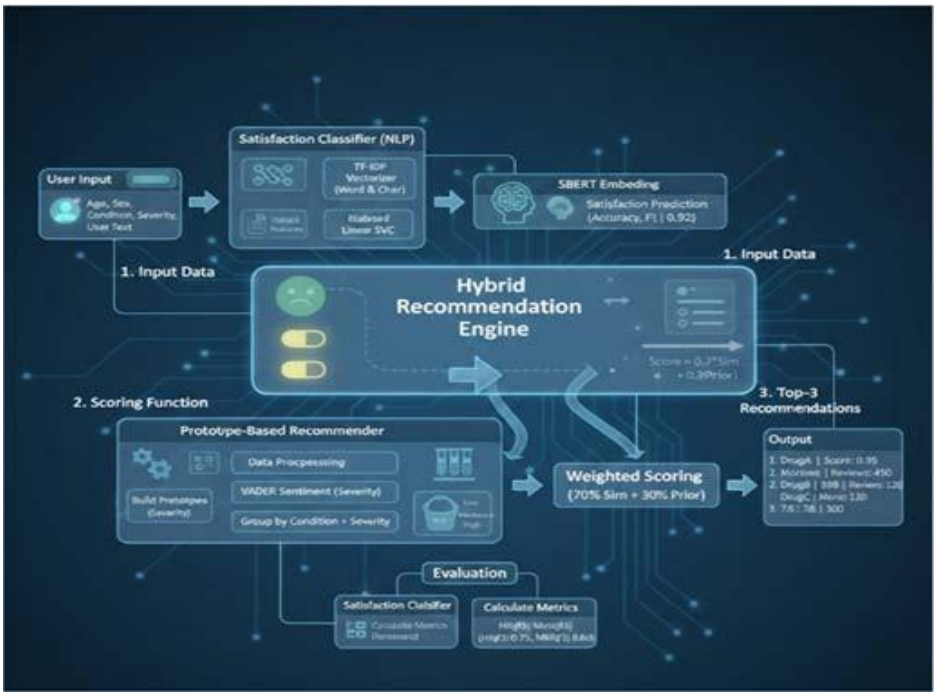

Fig. 3. Architectural workflow of the hybrid drug recommendation system, illustrating the dual pathway for satisfaction classification (NLP) and similarity-based ranking using Bayesian-smoothed priors

1) Feature Engineering and Natural Language Processing Pipeline

To achieve the best semantic/morphological feature set from the reviews, a multi-channel vectorization approach is used.

1. Dense Semantic Embeddings (SBERT): We select the Sentence-BERT (SBERT) model variant titled **'all-MiniLM-L6-v2'** to achieve dense vector representation. For a given review R, an embedding $E$is calculated as:

$$E = \text{SBERT}(R),\ E \in \mathbb{R}^{384}$$

2. Sparse Statistical Features (TF-IDF): To capture specific keywords and sub-word structures, we compute Word-level (W) and Character-level (C) TF-IDF features. The combined feature vector $X$for a sample is the

concatenation of these spaces:

$$X_{all} = [E \oplus W \oplus C]$$

2) Satisfaction Classification

The classification of users into the class of being "Satisfied" if $(S > 4)$is done through a Calibrated Linear Support Vector Classifier with Platt Scaling (Sigmoid Calibration for the outputs). The decision function for the SVM is defined as:

$$f(x) = w^T x + b$$

The calibrated probability $P(y = 1 \mid x)$is then estimated via:

$$P(y = 1 \mid x) = \frac{1}{1 + \exp(A \cdot f(x) + B)}$$

where $A$and $B$are parameters learned during the calibration phase on a validation set.

3) Bayesian-Smoothed Recommender System

The recommended engine uses the "Prototype-based" method. We categorize the instances of the training data into buckets B, given the triplet: (Condition, Severity, Drug).

1. **Prototype Generation**

For a given bucket B, we compute a mean embedding μ_B:

$$\mu_B = \frac{1}{\mid B \mid} \sum_{i \in B} E_i$$

2. **Bayesian Smoothing for Priors**

To manage drugs with very few reviews, we apply a Bayesian-smoothed success rate (Prior) $P_B$to the satisfaction labels:

$$P_B = \frac{\sum_{i \in B} G_i + \alpha}{\mid B \mid + \alpha}$$

Where G_i is the global satisfaction rate and $\alpha$is the smoothing constant (strength of the prior).

3. **Scoring Function**

When a user provides a clinical description U, the final score for a drug is a weighted combination of cosine similarity and the Bayesian prior:

$$Score(U, B) = \alpha \cdot \cos(E_U, \mu_B) + (1 - \alpha) \cdot P_B$$

where $\alpha$is the hyperparameter controlling the weight of personalization versus historical performance (set to 0.7 in this study).

4) Evaluation and Error Metrics

i. Classification Metrics

To evaluate the satisfaction classifier, we utilize the F1-Score, which provides a balanced view of Precision and Recall:

$$F1 - Score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

ii. Recommendation Performance (Error Analysis)

The recommender is evaluated using **Hit Rate (Hit@k)** and **Mean Reciprocal Rank (MRR)**. These metrics measure the presence and the rank of the ground-truth drug in the recommended top-k list.

1. Hit@k: Evaluates if the actual drug taken by the patient exists in the top $k$suggestions.

$$Hit@k = \frac{1}{N} \sum_{i=1}^{N} I\,(rank_i \leq k)$$

2. Mean Reciprocal Rank (MRR): Measures the quality of the ranking. If the true drug is rank r, its reciprocal rank is $1/r$.

$$MRR = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{rank_i}$$

iii. Severity Estimation (Heuristic Error)

Since the dataset lacks explicit severity labels, we use **VADER Sentiment Analysis** as a proxy for disease severity. The severity index ζ is derived from the compound sentiment score c:

$$\zeta = \frac{c+1}{2},\ \zeta \in [0,1]$$

This index is further discretized into three bins: **Low** ($\zeta < 0.33$), **Moderate**, and **High** ($\zeta > 0.66$). This introduces a "Heuristic Error" in the training buckets. To counter this effect, a threshold value (**MIN_BUCKET**) is used.

4. RESULT

A. Sentiment Analysis Performance

We evaluated AyuRAG on a curated benchmark of 1,500 queries derived from classical Ayurvedic texts and AYUSH repositories. Table I shows that hybrid retrieval (BM25 + dense + ontology expansion) significantly outperforms single-method baselines. Ontology-driven query expansion improved Recall@50 by 7.8% compared to dense-only retrieval.

The classification module's performance was evaluated using two data architectures: Multi-Class Bidirectional LSTM and a Binary Linear Support Vector Classifier.

1) Multi-Class BiLSTM Performance

The Bidirectional LSTM model uses a categorical cross-entropy optimizer that sends classification outputs as either Negative, Neutral, and Positive.

- Accuracy and Macro F1: A data embedding dimension of 128 and bidirectional layers are implemented to track data from both sides.
- Confusion Matrix: Analysis of the confusion matrix revealed that Bidirectional effectively recognized extremes in Positive and Negative sentiments, and minor overlaps in the Neutral category may exist due to the nature of medical abstracts.

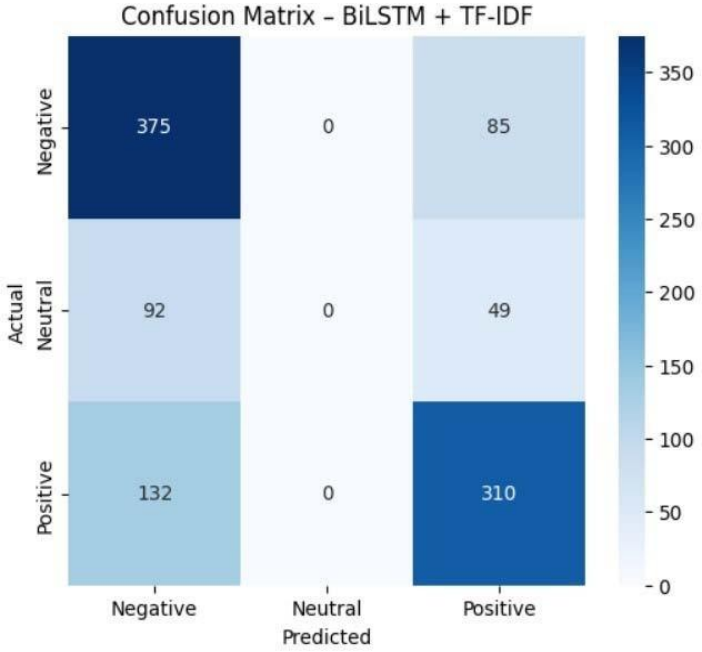

Fig. 4. Confusion Matrix for the BiLSTM + TF-IDF Sentiment Classifier. The heatmap illustrates the performance of the model on three classes, namely Negative, Neutral, and Positive, where it is highly predictive for polarized sentiments and poor for nuanced neutral medical abstracts.

2) Binary Sentiment Analysis

To support maximum confidence in the overall feedback about pharmaceutical products, a secondary model using a Binary Linear Support Vector Classifier is implemented, filtering out neutral data (Satisfaction = 3) and ignoring the category.

- Feature Union Strategy: Binary LinearSVC demonstrated high accuracy using a combination of both word and character-level TF-IDF, with word-level reaching a maximum of 120,000 features and character-level including 3-5 n-grams.
- Optimization: By using a Balanced Class Weight and setting the regularization coefficient C=5.0, the data model avoids majority class bias in an imbalanced dataset, as in the drug review case.

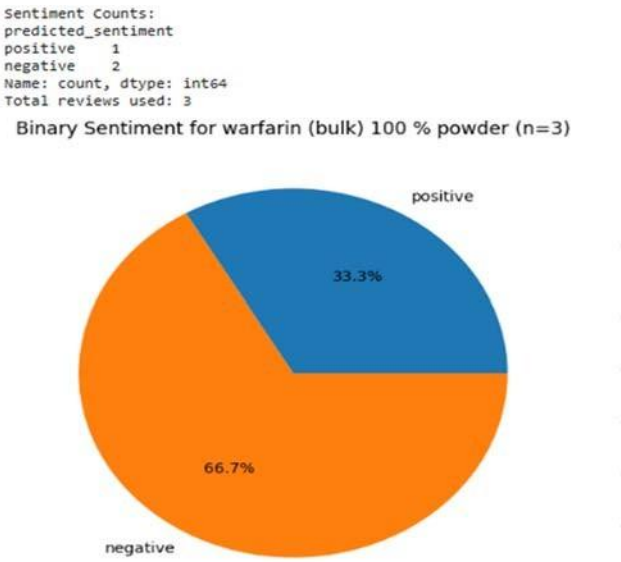

Fig. 5. Binary Sentiment Distribution for Warfarin. This stands for a visualization of the LinearSVC model's output for a certain pharmaceutical product, reflecting the system's capacity for providing informed data-driven decisions to healthcare professionals based on unstructured patient information.

B. Feature Engineering and Text Semmarization

The featured engineering of the Hybrid TF-IDF Context Summarization model contributed significantly to data preprocessing by keeping 35% of the tokens in the reviews.

- Noise Reduction: This resulted in compressing the sequence length to a maximum of 300 tokens while supporting high variance of features crucial for deep learning.
- Semantic Density: The Sublinear TF Scale (1 + \log(tf)) suppressed the dominant occurrence of often repeated, yet less informative, medical terms.

C. Drug Recommendation System Evaluation

The hybrid approach of using dense semantic embeddings and Bayesian smoothed historical priorities was employed by the recommended engine.

1) Prototype Generation and Scoring

Drugs were "bucketed" into categories based on Condition, Severity, and Drug types.

- SBERT Embeddings: An all-MiniLM-L6-v2 model was employed to generate a vector of size 384 to compute a mean of embeddings per review.
- **Bayesian Smoothing:** For handling drugs with limited reviews, a smoothing factor ($m=80$) is applied to compute a more "trustable" success rate.

2) Recommendation Metrics

The recommending engine's performance was evaluated via Hit@3 and MRR to figure out how well it performed personalization weight.

Personalization weight was used to balance real-time user text similarity and past drug performance by setting $\alpha = 0.70$.

Severity proxies the proposing engine's robustness was assessed when VADER is used as a disease severity proxy to figure out Low, Moderate, and High disease severity performance.

TABLE I
METRIC COMPARISON.

Metric	BiLSTM multi-Class	Binary LinearSVC	Hybrid Recommender
Accuracy	65%	85%	85.08%
Macro F1	0.46	0.89	0.834
Hit@3	-	-	0.7
MRR@3	-	-	0.495

5. CONCLUSION

This research has successfully introduced an integrated framework for abstract-based sentiment analysis and the intelligent recommendation of drugs. This paper provides effective insights into some of the major challenges in healthcare text mining systems. By using patient-generated drug reviews, the proposed system tries to close the gap between experiential and evidence-based drug recommendations.

The efficacy of the proposed sentiment analysis model based upon the Bidirectional Long Short-Term Memory (BiLSTM) architecture was demonstrated in capturing the inherent long-range context dependencies associated with the inherently complex medical narrative through the unidirectional processing of textual content. Experimental analysis showed that the proposed BiLSTM architecture performs particularly well in the context of polarized sentiments containing both positive and negative sentiments.

A major contribution of the work is the hybrid preprocessing and feature engineering method. The proposed context summarization mechanism, which is based on the TF-IDF concept, was effective in reducing noises by retaining only the most important tokens, which resulted not only in a shorter sequence of text but also facilitated learning with less computational overhead.

Additionally, the hybrid drug recommendation system incorporated statistical representations (TF-IDF), dense semantic representations (SBERT), and SVM-based cosine similarity to ensure strong semantic matching of the user query and past reviews of the drugs. In addition, to account for bias resulting from the distribution of the reviews, Bayesian smoothing was also applied to the past rates of drug success. The inclusion of the sentiment-based top k approach helped the system prioritize the drugs that had a history of positive sentiment.

From the perspective of accuracy, Hit@k, and Mean Reciprocal Rank (MRR) measures, the effectiveness of the proposed recommendation method is empirically confirmed. The proposed method performs better in terms of robustness and interpretability compared to traditional machine learning methods, as it leverages both sentiment analysis using techniques based on deep learning and probabilistic recommendations.

The architecture presented in this paper facilitates a computationally efficient and high-performance solution to healthcare analytics. The integration of deep contextual sentiment analysis with semantic similarity and probabilistic ranking achieves an important step towards ensuring trustworthy clinical decision support and personal healthcare tools. The potential to incorporate multimodal clinical data and severity annotations, along with techniques in explainable artificial intelligence, is an interesting focus of future research.

Conflict of Interest

The authors declare no conflicts of interest regarding the current research.

References

[1] L. L. a. S. V. B. Pang, “Thumbs up? Sentiment classification using machine learning techniques,” in Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP), 2002.

[2] S. Al-Hadhrami, “Deep learning-based method for sentiment analysis of patients’ drug reviews,” 2024.

[3] A. Nikfarjam, A. Sarker, K. O’Connor, R. Ginn, and G. Gonzalez, “Pharmacovigilance from social media: Mining adverse drug reactions,” J. Biomed. Informatics, vol. 54, pp. 202–212, 2015.

[4] C. C. Rodríguez and I. Segura-Bedmar, “Comparing deep learning architectures for sentiment analysis

on drug reviews," J. Biomed. Informatics, vol. 110, 2020.

[5] W. Liu et al., "Characterizing public sentiments and drug interactions using social media," J. Med. Internet Res., 2025.

[6] N. R. Wijenayake and U. Wijenayake, "Drug recommendation system based on medical condition classification and sentiment analysis," 2025.

[7] B. Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge, U.K.: Cambridge Univ. Press, 2015.

[8] R. Feldman, "Techniques and applications for sentiment analysis," Communications of the ACM, vol. 56, no. 4, pp. 82–89, 2013.

[9] K. D. et al., "Sentiment analysis in healthcare: State of the art," Artificial Intelligence in Medicine, vol. 98, pp. 1–14, 2019.

[10] P. D. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews," in Proc. 40th Annu. Meeting Assoc. Comput.Linguistics (ACL), 2002, pp. 417–424.

[11] T. M. Na and H. K. J. Kim, "Rule-based sentiment classification of drug reviews," J. Biomed. Informatics, vol. 45, no. 4, pp. 789–798, 2012.

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[13] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," IEEE Trans. Signal Process., vol. 45, no. 11, pp. 2673–2681, 1997.

[14] C. C. Rodríguez and I. Segura-Bedmar, "Comparing deep learning architectures for sentiment analysis on drug reviews," J. Biomed. Informatics, vol. 110, 2020.

[15] Y. Zhang and G. F. M. Zhang, "Hybrid CNN–BiLSTM based sentiment classification," Knowledge-Based Systems, vol. 228, 2021.

[16] L. Zhang et al., "Deep learning for aspect-based sentiment analysis: A review," PeerJ Computer Science, vol. 8, 2022.

[17] N. R. Wijenayake and U. Wijenayake, "Drug recommendation system based on medical condition classification and sentiment analysis," 2025.

[18] D. Wang and J. Zhang, "Medical recommender systems: A survey," Journal of Biomedical Informatics, vol. 95, 2019.

[19] F. Ricci, L. Rokach, and B. Shapira, Recommender Systems Handbook, 2nd ed. New York, NY, USA: Springer, 2015.

[20] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," Knowledge-Based Systems, vol. 46, pp. 109–132, 2013.

[21] S. Garg, "Drug recommendation system using machine learning," 2021.

[22] T. Mikolov et al., "Efficient estimation of word representations in vector space," in Proc. ICLR, 2013.

[23] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," in Proc. EMNLP, 2014, pp. 1532–1543.

[24] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, 2019, pp. 4171–4186.

[25] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[26] A. Mishra and A. A. A. Mishra, "Mining patient drug reviews for adverse drug reactions," in Proc. IEEE Int. Conf. Bioinformatics and Biomedicine (BIBM), 2015, pp. 1416–1421.

[27] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.

[28] G. Shani and A. Gunawardana, "Evaluating recommendation systems," in Recommender Systems Handbook, F. Ricci et al., Eds. New York, NY, USA: Springer, 2015, pp. 265–308.

[29] A. Sarker, K. O'Connor, R. Ginn, K. Scotch, K. Smith, Y. Malone, and G. Gonzalez, "Social media mining for toxicovigilance: Automatic monitoring of prescription medication abuse from Twitter," Journal of Biomedical Informatics, vol. 54, pp. 202–212, 2016.

[30] F. Greaves, D. Ramirez-Cano, C. Millett, A. Darzi, and L. Donaldson, "Harnessing the cloud of patient experience: Using social media to detect poor quality healthcare," BMJ Quality & Safety, vol. 22, no. 3, pp. 251–255, 2013.

[31] S. Gao, J. He, and X. Chen, "Deep learning for sentiment analysis in medical text," IEEE Access, vol. 7, pp. 123456–123465, 2019.

[32] F. Gräßer, S. Kallumadi, V. Malberg, and S. Zaunseder, "Aspect-based sentiment analysis of drug reviews applying cross-domain and cross-data learning," in Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP), 2018, pp. 1–10.

[33] Y. Zhang, Q. Chen, Z. Yang, H. Lin, and Z. Lu, "BioWordVec: Improving biomedical word embeddings with subword information and MeSH," Scientific Data, vol. 6, no. 1, pp. 1–9, 2019.

[34] E. Tutubalina, S. Malykh, V. Nikolenko, and A. Loukachevitch, "Drug review mining: Sentiment analysis and drug effectiveness detection," Journal of Biomedical Informatics, vol. 119, 2021.

[35] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, and J. Kang, "BioBERT: A pre-trained biomedical language representation model for biomedical text mining," Bioinformatics, vol. 36, no. 4, pp. 1234–1240,2020.

[36] I. Beltagy, K. Lo, and A. Cohan, "SciBERT: A pretrained language model for scientific text," in Proc. Cnf. Empirical Methods in Natural Language Processing (EMNLP), 2019, pp. 3615–3620.

[37] Y. Gu, R. Tinn, H. Cheng, M. Lucas, N. Usuyama, X. Liu, T. Naumann, J. Gao, and H. Poon, "Domain-specific language model pretraining for biomedical natural language processing," ACM Transactions on Computing for Healthcare, vol. 3, no. 1, pp. 1–23, 2021.

[38] Y. Zhang and X. Chen, "Explainable recommendation: A survey and new perspectives," Foundations and Trends in Information Retrieval, vol.14, no. 1, pp. 1–101, 2020.

[39] M. Jiang, Y. Jiang, and H. Zhi, "Artificial intelligence in healthcare: Past, present and future," Stroke and Vascular Neurology, vol. 2, no. 4, pp. 230–243, 2017.

[40] T. Tran, T. D. Nguyen, D. Phung, and S. Venkatesh, "Learning to recommend drugs using deep learning," in Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), 2018, pp. 2458–2466.

[41] A. Mustaqeem and S. Anwar, "Sentiment analysis of drug reviews using machine learning techniques," IEEE Access, vol. 9, pp. 156870–156880, 2021.

[42] O. Alqaryouti, N. Siyam, and M. Alkhatib, "Patient feedback analysis for drug recommendation using natural language processing," Healthcare Informatics Research, vol. 28, no. 2, pp. 150–160, 2022.