Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

Patient Emotion and Sentiment Analysis Using Deep Learning

Ajay Rastogi[1], Ravendra Singh[2], Mohammad Zubair Khan[3], Ibrahim Aljubayri[4], Fahad M Ghabban[5]

[1,2]Department of Computer Science and Information Technology, MJP Rohilkhand University, Bareilly, India.
[3,4]Department of Computer Science and Information, Taibah University Madinah, Saudi Arabia.
[5]College of Computer Science and Engineering. Information System Department, Taibah University Madinah, Saudi Arabia.

E-mail: [1]ajayrastogimbd@gmail.com, [2]r.singh@mjpru.ac.in, [3]zubair.762001@gmail.com, [4]ijubayri@taibahu.edu.sa, [5]fghaban@taibahu.edu.sa

Abstract

With the advancement in information technology and social media, there is a need for a patient sentiment-based healthcare system (HCS) to gauge its actual performance. Mining patient sentiment can provide valuable information about the strengths and weaknesses of these HCS and also help in their improvement. Sentiment analysis can help transform existing HCS into better and more efficient systems. Traditional sentiment analysis models are inefficient at capturing contextual information and high-dimensional data. This study aims to develop a classification model that allows the administration to evaluate the effectiveness of the healthcare system. This article proposes a hybrid deep neural network model, DeepCLNet, using the whale optimization technique (WOT) to enhance sentiment classification accuracy and fine-grain emotion analysis of patients. This model incorporates an adaptive feature weighting strategy that dynamically adjusts the weights and provides contextual sentiment refinement to deal with medical text. Furthermore, to enhance the capabilities of the model, this study proposes another hybrid model, BEDeepCLNet, by replacing the DeepCLNet embedding layer with BERT embeddings. The proposed models achieved state-of-the-art accuracies of 96% and 98%, respectively. During the experiments, it was observed that the proposed model performed better than other existing deep learning models applied in this domain. We also compared the proposed model's performance with the best NLP transformers like BERT, RoBERTa, XLNet, and SWIN, etc. The proposed model utilizes fewer computing resources while offering better performance than the transformers.

Keywords: Analysis, Deep Learning, Whale Optimization Algorithm, Healthcare services, Natural Language Processing, Adaptive Feature Weighting, Contextual Sentiment Refinement.

1. Introduction

Understanding patient sentiment or feedback is essential in the modern healthcare environment for the improvement of medical services as well as patient satisfaction [1]. It is a challenging task due to the highly complex nature of the data. Millions of patients have accounts on different social media platforms from all over the world; patients can express their thoughts and opinions at any time on social media. Twitter is mostly used by researchers to

DOI: https://doi.org/10.36548/jiip.2025.3.001
Received: 03.06.2025, received in revised form: 02.07.2025, accepted: 16.07.2025, published: 25.07.2025

collect data [1][2]. Patients generally share their opinions about their health issues. They share issues related to the services provided by the HCS [3]; indirectly, they are expressing the real performance [4] of HCS. It is very difficult to collect and process such genuine reviews from patients. These reviews are mainly posted on Twitter.

Sentiment analysis (also known as opinion mining) is a branch of natural language processing (NLP) that focuses on identifying, extracting, and classifying opinions or emotions expressed in text [5][6]. SA helps many industries gain insights into product quality. SA can also help to improve the HCS. This will be beneficial in improving the quality of services for patients [7]. While figuring out feedback among patients, SA synthesizes the good and bad aspects of the services. This provides a real picture of HCS among patients as well as other stakeholders. SA is not only used in the service industry like healthcare but also in other industries for the betterment of production. The healthcare industry is the backbone of the country. It is the largest industry serving people to recover their physical and mental health. The government always wants to evaluate the performance of its HCS. Yet, they have many other regulatory bodies to evaluate; they also want to assess genuine patient feedback. There are two approaches that help to mine patient sentiments. One is traditional machine learning methods, in which one has to engineer the features carefully, such as SVM, RF, etc. The second is deep learning methods like CNN and RNN that have proven to be leading in extracting lower-level features. Deep learning has proven to provide human-like understanding [8][9] and performs well in language understanding and classification. This study proposes two hybrid models. The first one is the hybrid DeepCLNet. It is the combination of several layers of CNN and LSTM. The DeepCLNet layered architecture is represented by Fig. 1. This model not only classifies sentiments but is also capable of classifying the emotions (anxiety, fear, hope, or frustration) of patients. Hyperparameters were optimized through the whale optimization technique. The other hybrid model is BEDeepCLNet. It is also capable of classifying the sentiments as well as the emotions of patients.

For comparison, this study experimented with many baseline deep learning models as well as state-of-the-art transformers. Finally, the major contribution of this work can be summarized as follows-

- This study proposes two hybrid models, *DeepCLNet* and *BEDeepCLNet* for sentiment and emotion analysis (anxiety, fear, hope, or frustration) in healthcare.
- The novelty of this study is
    - to add an attention layer with two heads for sentiment classification as well as to add fine grained emotion classification.
    - to use a BERT embedding layer and attention layer with the proposed model and develop a new hybrid model BEDeepCLNet.
    - to use the whale optimization technique to fine-tune the hyper-parameters of both models.
- This study compares the proposed model with baseline deep learning models such as RNN, LSTM, GRU as well as the state-of-the-art transformer models such as BERT [10], RoBERTa [11], DistilBERT [12] , XLNet [13] and SWIN-transformer [14].

The paper follows this structure: Section II introduces related work. Section III highlights the methodology covering the theoretical concepts, data collection, data exploration and the general framework of the proposed model. Section 4 discusses the experimental setup, findings and discussions. Finally, Section 5 concludes the paper.

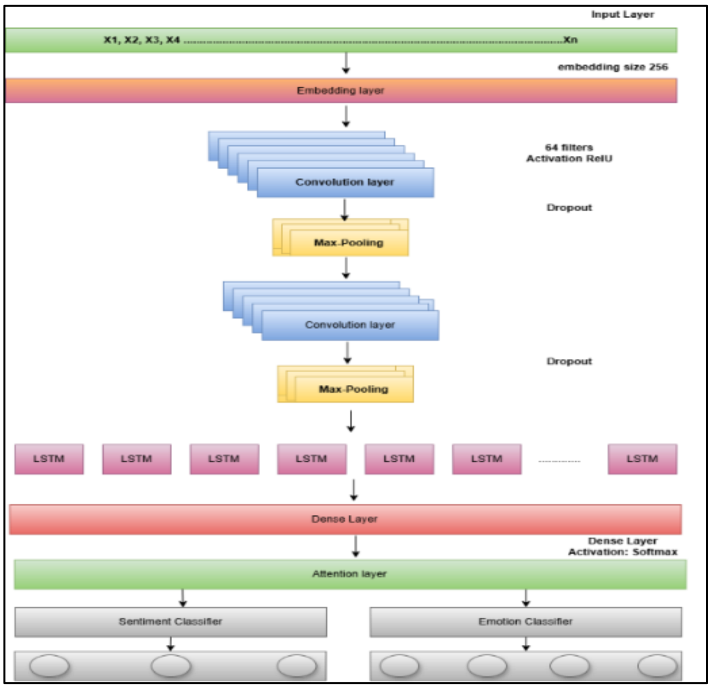


Figure 1. DeepCLNet Layered Architecture

2. Related Work

Opinions can be mined and categorized using Lexicon-based methods, machine learning, deep learning, and natural language processing. One of the simplest techniques for sentiment analysis is the lexicon-based approach. SentiWordNet is the most widely used lexicon, and SVM and Naïve Bayes are the most widely used machine learning techniques. According to the author [3], a lexicon-based approach should be used if time and data are limited; otherwise, a machine learning approach is advised. This is due to the fact that the machine learning approach necessitates a substantial amount of time for training as well as a large dataset. According to [15], natural language processing is the method that allows one to identify sentiments with human-like accuracy.

According to [16], opinion mining is performed using either a machine-learning-based approach or a rule-based approach. According to [2], Twitter is one of the most widely used microblogging tools by people. [5] used Twitter to analyze posts related to personal health information. [5] experimented with DT, SVM, and NB to classify the tweets. MaxEnt and SVM were proven to be the most effective in text classification in [1]. The authors used MaxEnt and NLP to predict the health status of patients from Twitter reviews. [17] examined six hundred forty-three internet comments that represent viewpoints on healthcare services in various nations. The author determined the frequency of healthcare-related subjects in the feedback. He also made an effort to categorize and rank the national HCS ratings of the respondents. This study advances the understanding of ML strategies and user-generated comments in practical ways. According to [18], NLP has become a powerful technique in recent years. The authors have identified two algorithms of NLP: NER (named entity recognition) and REM (relation extraction method) to obtain content information. [19] proposed a misinformation detection system for online information. The authors identified this type of misinformation spreading, whether intentionally or unintentionally, that provides fake information about healthcare prospects. [19] proposed CSM with a feature-based approach to identify misinformation within healthcare URLs. According to [20], opinions and facts are two different things that can be extracted from textual information. [21] designed an ensemble drift detection algorithm to detect changing behavior in client opinions. They used this algorithm to classify opinions on Amazon product reviews and IMDb movie review datasets and achieved more accurate results. According to [1], people's health status, whether good or bad, can be inferred from their messages on micro-blogging social media posts. It is also believed that the effectiveness or ineffectiveness of a patient's treatment is proportionally related to the response of the healthcare agency. A person posts positive feedback only when the agency's services are suitable for successful treatment. [22] proposed a pain detection system using sentiment analysis. The authors captured facial expressions using a camera and created the model by fusing statistical and deep learning-based techniques. [23] developed a rating mechanism based on the ABSA method. The authors considered four aspects: Doctors, Staff services, Hospital facilities, and Affordability. They collected over 30,000 online reviews of more than 500 hospitals. [23] provided ratings for the hospitals, which really helps patients select the right hospital. [4] suggested that healthcare professionals should remain connected with their patients. The authors explained that iPRO (Invisible Patient Reported Outcomes) represents the emotions that patients are unable to express to doctors. Online reviews are the only place where patients express this sentiment, as indicated in [4], by combining supervised machine learning with natural language processing. [7] proposed a patient opinion mining model that integrates data from social media and the real healthcare system. The authors used a transition-based approach to reconcile the differences between informal patient language and the formal medical language used in standard oncology. The proposed model employs a mixed dictionary and medical text dependencies, as well as idea transformation algorithms. Specifically, opinion classification models share neural network layers, allowing for a deeper understanding and application of related sentiment information encoded within conceptualized features in different contexts [7]. [24] proposed a weighted TF-IDF word embedding scheme with a deep learning CNN+LSTM technique, which performs well and achieves good accuracy in the domain of product reviews.According to [25], the deep learning method is able to learn from its multiple layers and tune itself. DL is capable of learning attributes of the data and generating fine results. Because DL has performed better in many domains due to its feature extraction ability, it is also used in text, opinion, and sentiment analysis. Hai Ha Do used several deep learning approaches. The authors advocated for deep learning techniques in aspect-level sentiment analysis. They used more than thirty-nine methods for this purpose. The authors summarized the best architecture for the classification task and also used pre-trained models to increase performance. They experimented with CNN, RNN, LSTM, Bi-LSTM, and GRU. The study involved examining linguistic elements and concept-based knowledge to improve performance [28]. Research studies use deep learning methods, including Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and Transformer-based architectures, to analyze sentiment in healthcare settings. The optimization process of hyperparameter tuning makes use of Genetic Algorithms (GA) and Particle Swarm Optimization (PSO). The hunting techniques of humpback whales have inspired WOA, which has become a successful method for both feature selection and optimization, enhancing model performance. The research intends to resolve two main issues in existing studies by implementing adaptive weighting for features together with context-based sentiment refinement.

Table 1. Summary of the Methods Used in Different Articles

Ref.	Method	Description	Application	Dataset	Results
[22]	SAS with the fusion of deep features & statistical features	captured facial emotions for sentiment prediction	Pain detection	shooting video of patients in McMaster University	SAS Accuracy: 83.71 on UNBC Shoulder Pain database and SAS Accuracy: 75.67% on D_2 database
[19]	CSM	Proposed CSM is better than the feature based method.	misinformation detection in healthcare	CoAID, FakeHealth, ReCOVery	Accuracy: 87.30 on CoAID, 86.30 ReCOVery, 85.26 FakeHealth datasets
[23]	ABSA, SentiWord Net	ABSA can aid customers in choosing services more wisely by enabling them to make more informed judgments.	Hospitals reviews	collected 30 hospitals data	-
[15]	CNN	combining different types of features from tweets to create tweet embedding	healthcare	Twitter	F1-Score:81.00%
[1]	MaxEnt + GIS algo	help healthcare providers to get in touch with continuously health condition	can predict health status of patient	Twitter	MaxEnt +GIS Accuracy:85
[7]	SVM, LSTM, Bi-LSTM	Combined given ML-DL with different combination of word vectorization-Word2Vec, Glove and Proposed a New Vectorization Scheme	Drug reaction	Twitter	Bi-LSTM + Proposed Vectorization Accuracy: 81%

[4]	multi-step sifting algorithm	capture iPRO, inform professionals and lead to more connected professionalized treatment	Healthcare	Twitter	-

3. Methodology

This study will employ a hybrid deep learning approach to analyze patient opinions, combining several neural network architectures to maximize the strengths of each model. The methodology consists of the following steps:

3.1 Data Collection and Preprocessing

For data collection, Kaggle, UCI, and GitHub were explored to find a suitable dataset. Patient reviews were required. The selected dataset is available on Kaggle [26], which contains tweets related to reviews of healthcare services. In addition to this, 25,000 Twitter posts were scraped using the Twitter API. This data was then merged with the Kaggle dataset (87,051 + 25,000). The final dataset contains 112,051 records. As the dataset did not have labels, it was annotated after preprocessing. NLTK was used for preprocessing, and sentiment annotation was performed using TextBlob [27], SentiWordNet [28], and VADER [29] lexicon techniques.

Table 2. Unique Label Count in Dataset

Label	Description	Count
Good	Indicate the efficiency of HCS	51,564
Bad	Indicate the deficiency of HCS	32,840
Neutral	Indicate the review is just related to the HCS.	27,647

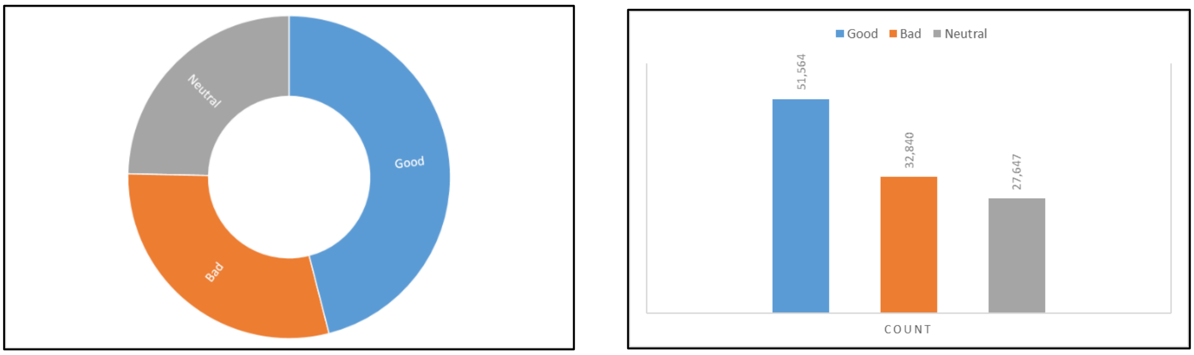


Figure 2. Data Distribution within Three Classes

The dataset is classified into three labels: good, bad, or neutral. A good review shows the effectiveness of an HCS. A bad review shows the deficiency of the HCS. A neutral review indicates that the review is just related to the HCS. Table 2 represents the dataset distribution across the three classes: Good, Bad, and Neutral. This is a highly imbalanced dataset. There are two techniques to balance the dataset: data augmentation or data removal. To maintain the integrity of the reviews, this study selects the removal method for balance. Text pre-processing steps are described in detail in the following section.

Text Pre-Processing Steps Applied on Dataset.

BEGIN

// *Step* 1: *Remove duplicate messages*

REMOVE duplicate messages from dataset

// *Step* 2: *Convert all data into lowercase*

FOR each message in dataset:

CONVERT message to lowercase

// *Step* 3: *Remove special characters from the data*

FOR each message in dataset:

REMOVE special characters like "#", "@", "!", "..." *from message*

// *Step* 4: *Remove numbers from the data*

FOR each message in dataset:

REMOVE numbers from message

// *Step* 5: *Remove extra white space*

FOR each message in dataset:

REMOVE extra white spaces from message

// *Step* 6: *Remove punctuation*

FOR each message in dataset:

REMOVE punctuation marks from message

// *Step* 7: *Remove Stop words*

FOR each message in dataset:

REMOVE stop words from message

// *Step* 8: *Perform stemming and lemmatization*

FOR each message in dataset:

$PERFORM\ stemming\ on\ message$

$PERFORM\ lemmatization\ on\ message$

$//\ Step\ 9\text{: } Annotation\ of\ dataset\ into\ three\ Classes - Positive, Negative, and\ Neutral$

$FOR\ each\ message\ in\ dataset\text{:}$

$ANNOTATE\ message\ as\ Positive, Negative, or\ Neutral\ based\ on\ sentiment$

$//\ Step\ 10\text{: } Perform\ Dataset\ balancing$

$BALANCE\ dataset\ to\ ensure\ equal\ distribution\ of\ Positive, Negative, and\ Neutral\ classes$

END

3.2 Data Exploration and Analysis

The data is explored using python language. Fig. 4 (a) shows the top words frequencies in the good reviews and (b) shows the top words frequencies in the bad reviews. Most of the top ten words are common and shows relatedness of the review of patient sentiment towards HCS. Many frequent words are same as healthcare, feedback, review etc. The words thank, good, well, new, great, service indicates the quality of HCS in positive manner while leave, need, help, please, see, covid are the indicates the quality of HCS in negative manner. Negative words show the pressure and emotions of patients. The word-cloud is represented by Fig. 3.

Figure 3. Word-Cloud Representation of the Data

3.3 Data Splitting

Data is divided into 60% for training, 20% for validation and 20% for testing samples respectively.

3.4 Proposed Hybrid Method - DeepCLNet

The proposed DeepCLNet is a combination of one input layer, one text embedding layer, two convolution layers, one LSTM network, one dense layer, attention layer, one sentiment classification output layer and one emotion classification output layer. It classifies text into good, bad and neutral. The first part of DeepCLNet is CNN.

Convolution Layer: Conv1D is used for processing one-dimensional sequences, such as text. Filters: This parameter determines the number of filters (or kernels) that the layer will use. Each filter is responsible for learning different features from the input text. In this case, there are 32 filters.

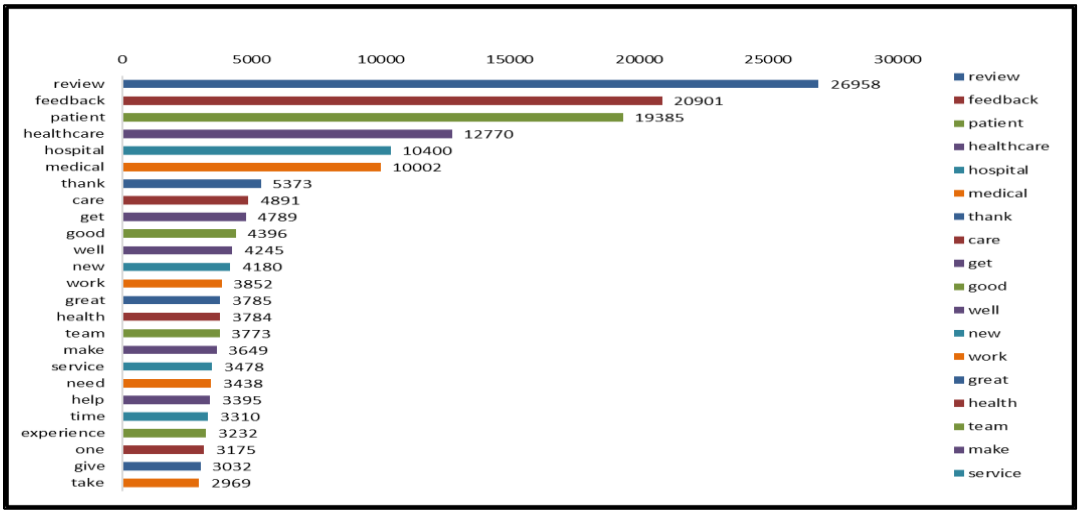


Figure 4 (a). Word frequencies in Good Reviews

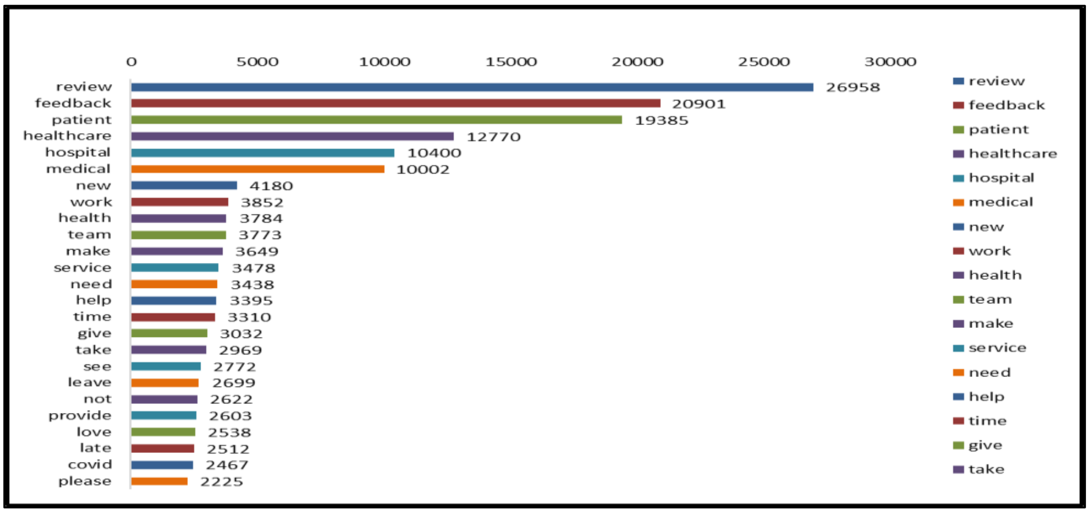


Figure 4 (b). Word Frequencies in Bad Reviews

Activation Function (ReLU): After producing the feature map, the ReLU activation function is utilized. It introduces nonlinearity into the network and is known for its simplicity and speed. It replaces negative values with zero, which accelerates training and reduces computation time.

$$Y_i = ReLU\,(\sum_{j=1}^{n} W_j * X_{i+j-1} + b) \tag{1}$$

MaxPooling1D is a pooling layer used to down-sample the spatial dimensions of the input sequence. It retains the most important information from the output of the convolution layer. This parameter sets the size of the pooling window and specifies how many adjacent values are considered in the pooling operation. In this case, the pooling window size is 2.

$$Z_i = Max_{j=1}^{2} * Y_{2i+j-2} \tag{2}$$

Above equation represents convolution operation (1) where W_j represents the convolutional kernel, X_{i+j-1} is the input element at position $i + j - 1$ and max pooling operation (2) where Y_{2i+j-2} is the element at position $2i + j - 2$ element. In the third layer, the convolution operation with an L2 regularizer (regularization strength of 0.01) is repeated, followed by the second max-pooling layer.

Then the output was passed to the LSTM layer. LSTM layer has 150 units. LSTM have the six gates. f_t represents forget gate, i_t represents input gate, o_t is output gate $\widetilde{C_t}$, C_t, and h_t are the cell and hidden states as given in following equations.

$$f_t = \sigma\left(W_f \cdot \left[h_{\{t-1\}}, x_t\right] + b_f\right) \tag{3}$$

$$i_t = \sigma\left(W_i \cdot \left[h_{\{t-1\}}, x_t\right] + b_i\right) \tag{4}$$

$$\widetilde{C_t} = \tanh\left(W_C \cdot \left[h_{\{t-1\}}, x_t\right] + b_C\right) \tag{5}$$

$$C_t = f_t * C_{\{t-1\}} + i_t * \widetilde{C_t} \tag{6}$$

$$o_t = \sigma\left(W_o \cdot \left[h_{\{t-1\}}, x_t\right] + b_o\right) \tag{7}$$

$$h_t = o_t * \tanh(C_t) \tag{8}$$

It processes the sequence output from the CNN layers and produce a hidden state. followed by dropout layer. The output of LSTM passed to the attention layer. The attention layer uses learnable weights W_a and bias b_a *to* compute attention scores

$$a_t = softmax(W_a h_t + b_a) \tag{9}$$

Context vector (weighted sum of hidden states):

$$A = \sum_{t=1}^{T/4} a_t h_t \tag{10}$$

The share attention output is them move two different sides one is for main sentiment classification head Y_i

$$Y_i = Softmax\ (W_{main} A + b_{main}) \tag{11}$$

other side is for emotion classification head Y_j

$$Y_j = Softmax\ (W_{fine} A + b_{fine}) \tag{12}$$

Above equation determines the main sentiment classification and emotion classification. Both are connected with the attention layer.

3.5 Feature Extraction

Trainable-Embedding (WE-1) as well as Word2Vec (WE-2) were used for feature extraction. Both feature extraction techniques in natural language processing (NLP) capture the context of the message. The Word Embedding layer creates vectors to represent words. Two hundred and fifty vectors were created for each word using the Word Embedding layer.

The combination of Conv-1D and Max-pooling is a common pattern in CNNs. Conv-1D helps to capture local patterns from input text while pooling reduces the dimensionality of the features. The CNN layer in the model applies common convolutional filters to the input text. Each filter is a small sliding window that moves across the sequence length of the input text data. Here how the convolution operation works in feature extraction. For each filter f_k *(where* $k \in (1, \ldots, \text{number of filters}))$ the convolution operation involves sliding the filter across the sequence length of the input text data. The input data is a sequence of the embedded word vectors represented as a 3D tensor x with shape *(samples, sequence length, embedding layer)*. The filter f_k is a 2D tensor with shape *(filter size, embedding size)*. The convolution operation between filter $\boldsymbol{f_k}$ and a segment of the input sequence can be mathematically represented as:

$$y_{i,k} = \sum_{j=1}^{filter\_size} (x_{i,j} * f_{k,j}) + b_k \tag{13}$$

Where $y_{i,k}$ is the output of the convolution for the k-th filter at position i, $x_{i,j}$ is the input sequence at position $i+j$, $f_{k,j}$ is the value of the k -th filter at position j, b_k is the bias term for the k -th filter. After the convolution, an activation function (such as ReLU) is applied to the output to introduce non-linearity. The CNN layer produces a 3D output tensor with shape *(samples, sequence length, filters).*

This output contains the activation levels of each filter for each input sequence and each position in the sequence. Filters in the CNN layer are designed to capture specific features from the input data. These can be *n-grams, phrases or other patterns* in the text. Observing how activation levels vary across different samples and sequence lengths can provide information about what features each filter is capturing and how consistently it behaves across different input samples.

3.6 Hybrid Bert Embedding with DeepCLNet – BEDeepCLNet

This hybrid model is blended with BERT embeddings and tokenization layers. This model has the BERT feature extraction layer at the top. The first layer is the input layer. Input text is then transferred to BERT embeddings and attention. This generates sequential output. This sequential output is then transferred to two convolutional networks. This will eventually reduce the dimensionality as well as the network complexity. Then the output is passed to the LSTM, followed by dropout and the attention head. The attention head derives two different classifiers: one is used for sentiment classification, and the other network is used for emotion classification. The pipeline of BEDeepCLNet is represented in Fig. 5.

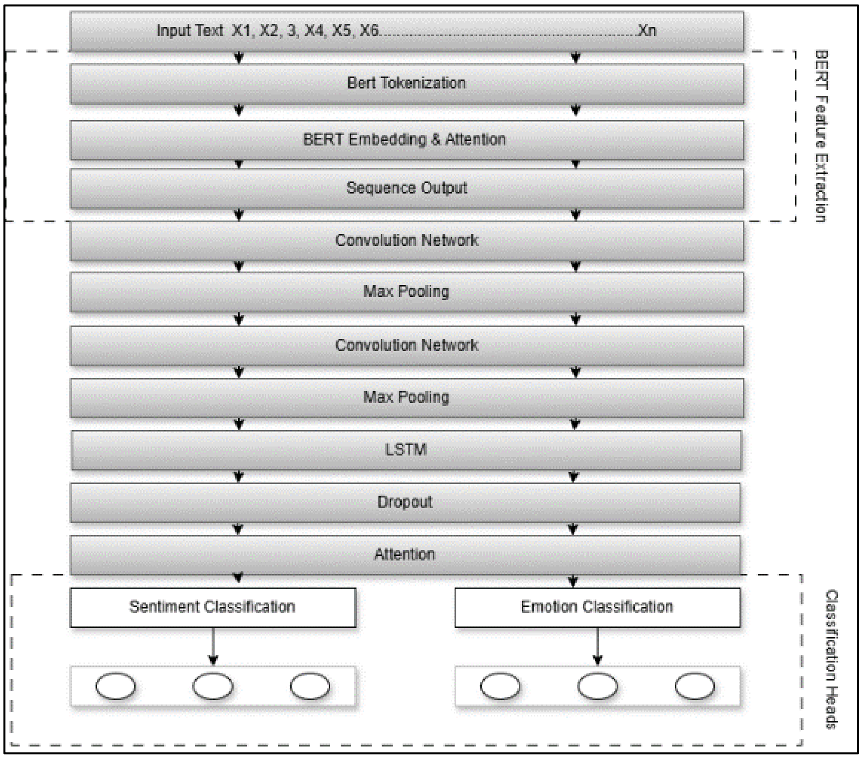


Figure 5. BEDeepCLNet Pipeline

3.7 Whale optimization technique

The whale optimization technique(WOA) [30].is used to optimize the hyper-parameters in our proposed deep learning model- DeepCLNet. It provides the best hyper-parameters to our model. WOA is a metaheuristic technique uses whale humpback hunting method. It is found useful in 7 different fields and 17 subfields of engineering domain. WOA is used to improve neural network architectures by hyper-parameter tuning and optimizing weight configuration and learning rate. The algorithm excels at balancing the exploration and exploitation phases, which is important in searching for global optima in complex, multi-dimensional spaces such as those found in deep learning models.

The three primary stages of the WOA operations are prey discovery, spiral bubble-net feeding, and prey encirclement. These processes align with optimization processes in which whale agents adjust their position with respect to a leader to iteratively refine the solution that is the best solution discovered so far. For problems such as feature selection, image classification and neural network training in deep learning, the algorithm's ability to avoid local minima through a dynamic balance between exploration (discovering new areas) and exploitation (improving existing solutions) makes it suitable.

3.8 Performance Measure and Evaluation Metrics

- The performance measurement is applied after training the model. The following measures are applied after generating the confusion matrix. The following equations are used for the performance measurement in terms of Accuracy, F1-score, Recall, and Precision. Accuracy measures the correct prediction ratio from total predictions. The F1-score provides a correct measure of accurate prediction if the dataset is imbalanced. In our scenario, there are four parameters required to calculate accuracy, precision, and recall: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).TP refers to correctly classified positive instances.

- TN refers to correctly classified negative instances.
- FP occurs when a negative instance is incorrectly classified as positive.
- FN occurs when a positive instance is incorrectly classified as negative.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{14}$$

$$Precision = \frac{TP}{TP+FP} \tag{15}$$

$$Recall = \frac{TP}{TP+FN} \tag{16}$$

$$F1 - Score = 2 * \frac{Precision*Recall}{Precision+Recall} \tag{17}$$

3.9 Experimental Setup

The proposed methods, *DeepCLNet* and *BEDeepCLNet*, were developed in a Google Colab notebook using the whale optimization technique. TensorFlow and Keras libraries were utilized for defining and training the neural network architecture. NLTK and Gensim libraries were employed for data preprocessing. These models classify text into three classes of sentiment as well as four classes of emotions, employing categorical cross-entropy to measure the loss. The following equation defines loss in a multiclass classification problem.

$$loss(l) = -\sum_{m=1}^{n} y_{i,m} \log(p_{i,m}) \tag{18}$$

This is the categorical cross-entropy loss, commonly used for multiclass classification problems. It measures how far the predicted class probabilities are from the actual labels. The loss is low when the predicted probability for the correct class is high, and it increases as the model becomes more confident in a wrong prediction. Here *n* is number of classes and *m* belongs to the class and *p* belongs to the probability. The learning rate is set to 0.001. 1818143928

4. Results and Discussion

The effectiveness of the suggested model DeepCLNet was assessed by testing it with well-known deep neural networks. The embedding layer has different embedding techniques such as WE-1 (custom trainable embeddings) and WE-2 (Word2Vec). We have taken RNN, LSTM, Bi-LSTM, and GRU models and used the given embeddings. We compare the results of these models with the proposed model DeepCLNet. The results are shown in Table 3. In Fig. 3 and Fig. 4, we can see the performance comparison between the baseline deep learning models and the proposed model. The difference in performance is plotted based on WE-1 and WE-2. DeepCLNet with WE-2 embedding demonstrates the highest performance, achieving an accuracy and F1-Score of 96%. The highest accuracies among the conventional models are shown by LSTM and Bi-LSTM with WE-2 embedding, suggesting that they are capable of capturing a respectable degree of context and semantics.

Table 3. Performance Metrics of Different Baseline Deep Learning Models

Model	WE-1				WE-2			
	Accuracy	F1	Recall	Precision	Accuracy	F1	Recall	Precision
RNN	0.7619	0.7583	0.7668	0.7501	0.7718	0.768	0.7629	0.7732
Bi-LSTM	0.8158	0.8095	0.8101	0.8091	0.8287	0.8257	0.8238	0.8258
GRU	0.7911	0.7883	0.7867	0.7901	0.8098	0.8146	0.8078	0.8105
LSTM	0.8127	0.8093	0.8101	0.8086	0.8278	0.8287	0.8358	0.8218
DeepCLNet	0.9364	0.9362	0.9367	0.9359	0.9687	0.9655	0.9678	0.9612

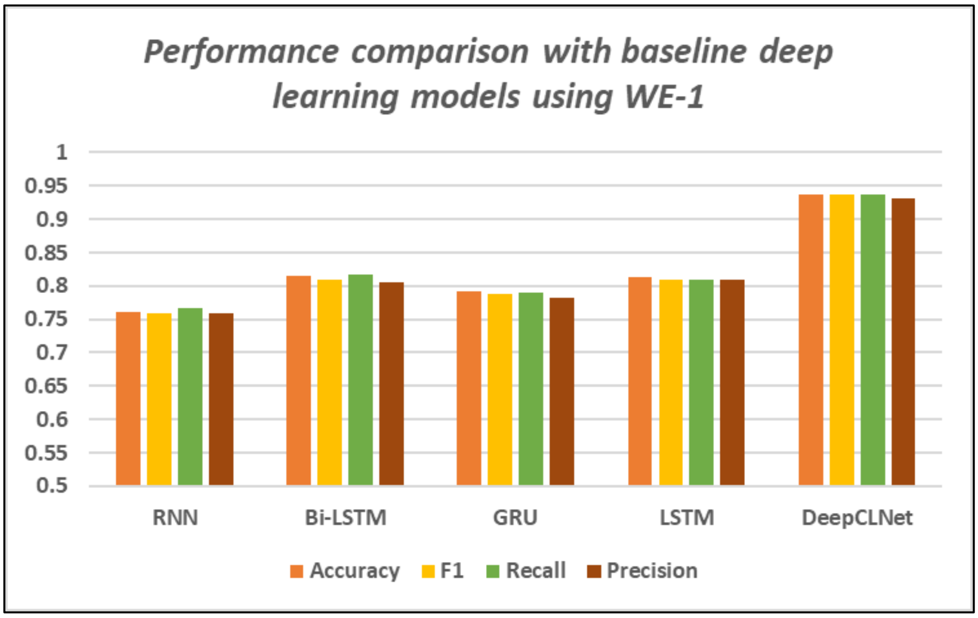


Figure 6. Comparison of DeepCLNet with Baseline Deep Learning Using WE-1

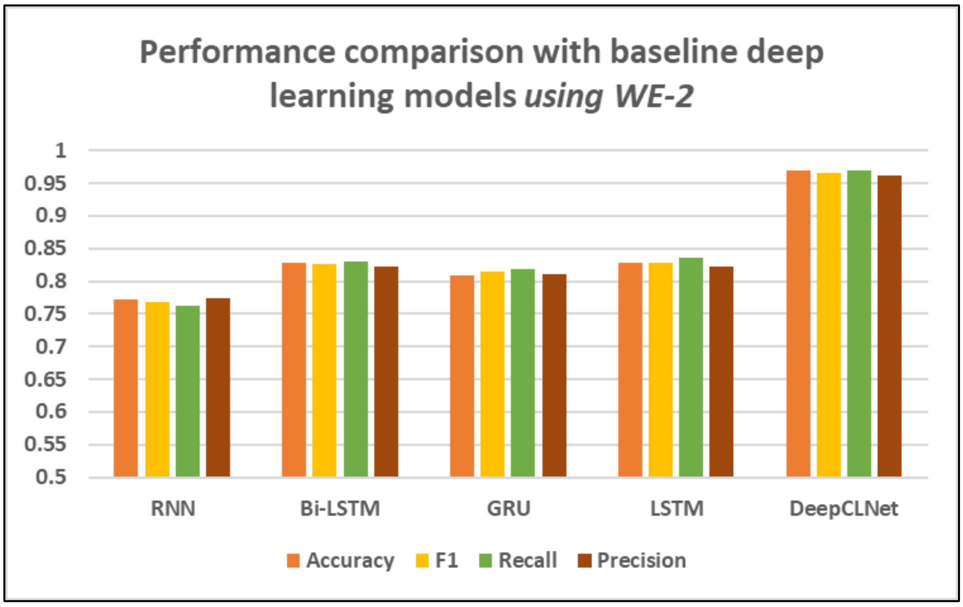
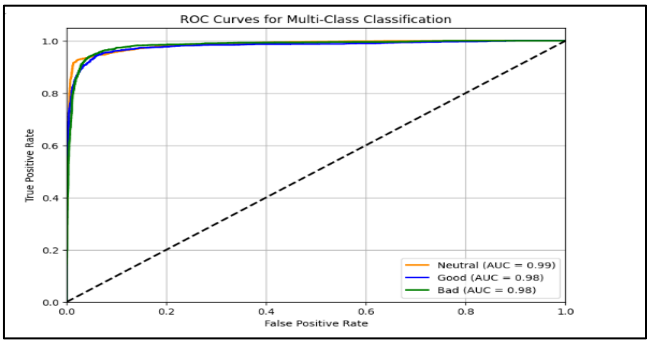
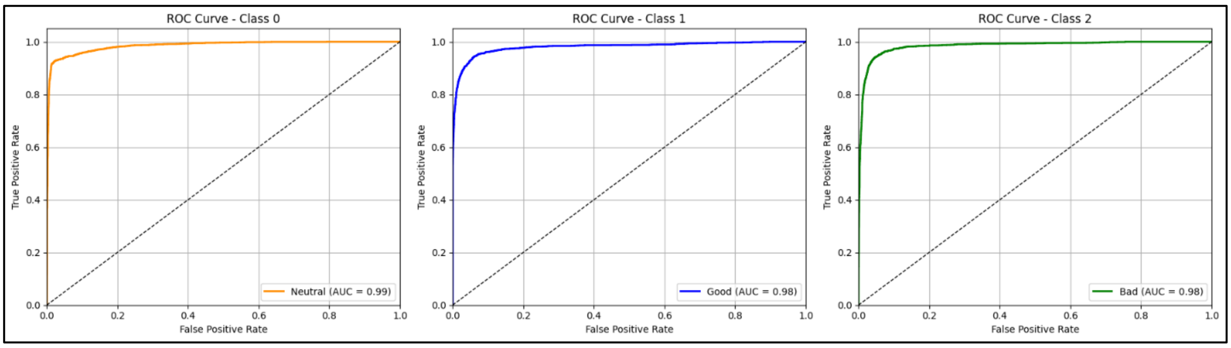


Figure 7. Comparison of DeepCLNet with Baseline Deep Learning Using WE-2



a)



b) c) d)

Figure 8. ROC Curves of a) Proposed DeepCLNet Model b) Neutral Reviews, c) Good Reviews, d) Bad Reviews

This curve displays the performance of proposed model on each class and effectiveness of classification.

Here, the ROC curve is ideal, but the accuracy is only slightly above 96.87%. This usually means the model is performing well in distinguishing between classes. The first reason is that ROC measures probability ranking, not exact labels. This means ROC-AUC checks how well the model ranks positive vs. negative cases, not whether it makes exact correct predictions. So, a model can have a high AUC (distinguishing well) but still misclassify some borderline cases, lowering accuracy. In multi-class problems (like three sentiment and four emotion classes), AUC might be computed per class or in a one-vs-rest fashion. High AUC for each class doesn't always translate into high overall accuracy.

Table 4. Comparison of Transformers Accuracy with Proposed Model

Transformers	Accuracy
BERT	0.9521
RoBERTa	0.9668
DistilBERT	0.9582
XLNet	0.9612
SWIN	0.9687
BEDeepCLNet	0.9821

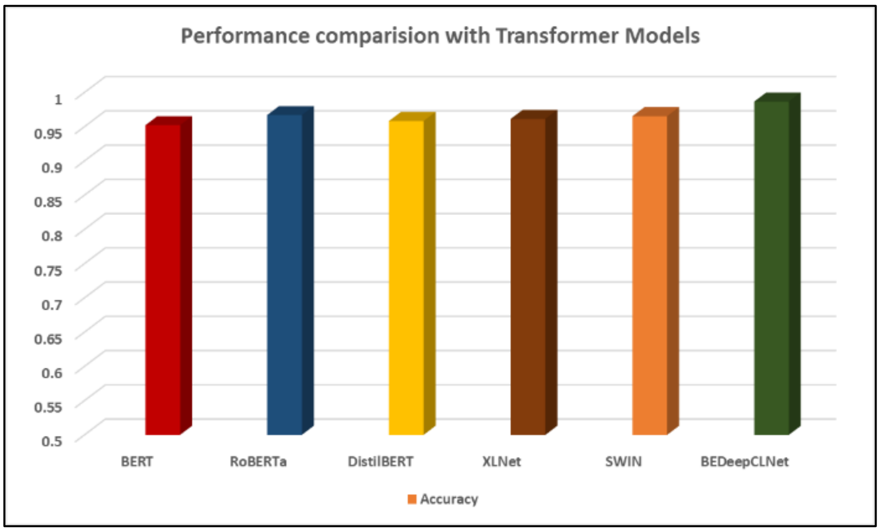


Figure 9. Comparison of Different Transformers Models with Proposed Model BEDeepCLNet

Table 5 shows the comparison between the proposed approach and the previous approaches. This table includes the methods, descriptions of the methods, applications, datasets, and results. It is clear from the results produced by the proposed models that they are unmatched, i.e., better than the previous work. The proposed approaches can be used for the detailed analysis of patient sentiment as well as the emotions of the patients, which are not addressed in any previous approaches.

Table 5. Comparison with Previous Studies

Ref.	Method	Results
[15]	CNN	F1-Score:81.00%
[7]	SVM, LSTM, Bi-LSTM	Bi-LSTM + Proposed Vectorization Accuracy: 81%
Proposed work	Hybrid models DeepCLNet and BEDeepCLNet	DeepCLNet acc:96%; BEDeepCLNet acc:98%

5. Future Enhancement

The hybrid architecture DeepCLNet and BEDeepCLNet, both combining CNN and LSTM layers, increase the number of parameters, leading to higher computational costs and longer training times, especially with large datasets. Future work will focus on developing domain-specific lexicons. These lexicons will be used to train the models for more accurate classification. Furthermore, different optimization techniques could be effectively used with deep learning hybrid models to gain more performance with fewer computational resources.

6. Conclusion

Patient sentiment reflects their health condition and directly impacts quality of life. Analyzing opinions shared on social media can help evaluate the performance and satisfaction levels of healthcare systems (HCS). This study introduces two hybrid deep learning models: *DeepCLNet* and *BEDeepCLNet*. DeepCLNet combines CNN, LSTM, and ANN layers and is optimized using the Whale Optimization Technique for efficient sentiment and emotion classification. *BEDeepCLNet* enhances *DeepCLNet* by incorporating BERT embeddings and tokenization for better contextual understanding.

DeepCLNet trains faster, while *BEDeepCLNet* achieves about 2% higher accuracy. Experimental results show that *DeepCLNet* outperforms traditional models such as RNN, LSTM, Bi-LSTM, and GRU. Two feature extraction methods were used: trainable word embeddings (WE-1) and Word2Vec (WE-2), with WE-2 delivering superior results. *BEDeepCLNet* also outperforms leading transformer-based models like BERT, RoBERTa, DistilBERT, XLNet, and SWIN-Transformer. Both models are resource-efficient and suitable for real-time sentiment analysis in healthcare applications.

The hybrid architecture *DeepCLNet* and *BEDeepCLNet,* both combining CNN and LSTM layers, increases the number of parameters, leading to higher computational cost and longer training times, especially with large datasets. In future work will focus, on developing domain-specific lexicon. These lexicons will be used to train the models for more accurate classification. Further different optimization techniques could be effectively used with deep learning hybrid models to gain more performance with fewer computational resources.

References

[1] Htet, Hein, Soe Soe Khaing, and Yi Yi Myint. "Tweets sentiment analysis for healthcare on big data processing and IoT architecture using maximum entropy classifier." In International Conference on Big Data Analysis and Deep Learning Applications, Singapore: Springer Singapore, 2018, 28-38.

[2] Alsaeedi, Abdullah, and Mohammad Zubair Khan. "A study on sentiment analysis techniques of Twitter data." International Journal of Advanced Computer Science and Applications 10, no. 2 (2019).

[3] RamyaSri, V. I. S., Ch Niharika, K. Maneesh, and Mohammed Ismail. "Sentiment analysis of patients' opinions in healthcare using lexicon-based method." International Journal of Engineering and Advanced Technology 9, no. 1 (2019): 6977-6981.

[4] Clark, Eric M., Ted James, Chris A. Jones, Amulya Alapati, Promise Ukandu, Christopher M. Danforth, and Peter Sheridan Dodds. "A sentiment analysis of breast cancer treatment experiences and healthcare perceptions across twitter." arXiv preprint arXiv:1805.09959 (2018).

[5] Saranya, S., and G. Usha. "A Machine Learning-Based Technique with IntelligentWordNet Lemmatize for Twitter Sentiment Analysis." Intelligent Automation & Soft Computing 36, no. 1 (2023).

[6] Sureja, Nitesh, Nandini Chaudhari, Priyanka Patel, Jalpa Bhatt, Tushar Desai, and Vruti Parikh. "Hyper-tuned Swarm Intelligence Machine Learning-based Sentiment Analysis of Social Media." Engineering, Technology & Applied Science Research 14, no. 4 (2024): 15415-15421.

[7] Grisstte, Hanane, and ELhabib Nfaoui. "Daily life patients sentiment analysis model based on well-encoded embedding vocabulary for related-medication text." In Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2019, 921-928.

[8] Abedin, Afia Fairoose, Amirul Islam Al Mamun, Rownak Jahan Nowrin, Amitabha Chakrabarty, Moin Mostakim, and Sudip Kumar Naskar. "A deep learning approach to integrate human-level understanding in a chatbot." arXiv preprint arXiv:2201.02735 (2021).

[9] Irwansyah, Edy, Alexander AS Gunawan, Hady Pranoto, Fabian Surya Pramudya, and Lucky Fakhriadi. "Deep Learning with Semantic Segmentation Approach for Building Rooftop Mapping in Urban Irregular Housing Complexes." Engineering, Technology & Applied Science Research 15, no. 2 (2025): 20580-20587.

[10] Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." In Proceedings of the 2019 conference of the North American chapter of the association for

computational linguistics: human language technologies, volume 1 (long and short papers), 2019, 4171-4186.

[11] Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. "Roberta: A robustly optimized bert pretraining approach." arXiv preprint arXiv:1907.11692 (2019).

[12] Dogra, Varun, Aman Singh, Sahil Verma, Kavita, N. Z. Jhanjhi, and M. N. Talib. "Analyzing distilbert for sentiment classification of banking financial news." In Intelligent Computing and Innovation on Data Science: Proceedings of ICTIDS 2021, Singapore: Springer Nature Singapore, 2021, 501-510.

[13] Yang, Zhilin, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R. Salakhutdinov, and Quoc V. Le. "Xlnet: Generalized autoregressive pretraining for language understanding." Advances in neural information processing systems 32 (2019).

[14] Agilandeeswari, Loganathan, and S. Divya Meena. "SWIN transformer based contrastive self-supervised learning for animal detection and classification." Multimedia Tools and Applications 82, no. 7 (2023): 10445-10470.

[15] Phan, Huyen Trang, Van Cuong Tran, Ngoc Thanh Nguyen, and Dosam Hwang. "Improving the performance of sentiment analysis of tweets containing fuzzy sentiment using the feature ensemble model." Ieee Access 8 (2020): 14630-14641.

[16] Rahab, Hichem, Hichem Haouassi, and Abdelkader Laouid. "Rule-based Arabic sentiment analysis using binary equilibrium optimization algorithm." Arabian Journal for Science and Engineering 48, no. 2 (2023): 2359-2374.

[17] Ruelens, Anna. "Analyzing user-generated content using natural language processing: a case study of public satisfaction with healthcare systems." Journal of computational social science 5, no. 1 (2022): 731-749.

[18] Khurana, Diksha, Aditya Koli, Kiran Khatter, and Sukhdev Singh. "Natural language processing: state of the art, current trends and challenges." Multimedia tools and applications 82, no. 3 (2023): 3713-3744.

[19] Barve, Yashoda, and Jatinderkumar R. Saini. "Detecting and classifying online health misinformation with 'Content Similarity Measure (CSM)'algorithm: an automated fact-checking-based approach." The Journal of Supercomputing 79, no. 8 (2023): 9127-9156.

[20] Ahmad, Pir Noman, Yuanchao Liu, Gauhar Ali, Mudasir Ahmad Wani, and Mohammed ElAffendi. "Robust benchmark for propagandist text detection and mining high-quality data." Mathematics 11, no. 12 (2023): 2668.

[21] Kumar, Sanjeev, Ravendra Singh, Mohammad Zubair Khan, and Abdulfattah Noorwali. "Design of adaptive ensemble classifier for online sentiment analysis and opinion mining." PeerJ Computer Science 7 (2021): e660.

[22] Ghosh, Anay, Saiyed Umer, Muhammad Khurram Khan, Ranjeet Kumar Rout, and Bibhas Chandra Dhara. "Smart sentiment analysis system for pain detection using cutting edge techniques in a smart healthcare framework." Cluster Computing 26, no. 1 (2023): 119-135.

[23] Bansal, Ankita, and Niranjan Kumar. "Aspect-based sentiment analysis using attribute extraction of hospital reviews." New Generation Computing 40, no. 4 (2022): 941-960.

[24] Onan, Aytuğ. "Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks." Concurrency and computation: Practice and experience 33, no. 23 (2021): e5909.

[25] Zhang, Lei, Shuai Wang, and Bing Liu. "Deep learning for sentiment analysis: A survey." Wiley interdisciplinary reviews: data mining and knowledge discovery 8, no. 4 (2018): e1253.

[26] "Healthcare_Dataset." [Online]. Available: https://www.kaggle.com/datasets/devkiagarwal/healthcare-dataset#

[27] Lorla dan Steven, "TextBlob Documentation Release 0.16.0," TextBlob, 2020, [Online]. Available: https://textblob.readthedocs.io/en/dev/

[28] Husnain, Mujtaba, Malik Muhammad Saad Missen, Nadeem Akhtar, Mickaël Coustaty, Shahzad Mumtaz, and VB Surya Prasath. "A systematic study on the role of SentiWordNet in opinion mining." Frontiers of Computer Science 15, no. 4 (2021): 154614.

[29] Marutho, Dhendra, and Supriadi Rustad. "Sentiment Analysis Optimization Using Vader Lexicon on Machine Learning Approach." In 2022 international Seminar on intelligent Technology and its applications (ISITIA), IEEE, 2022, 98-103.

[30] Okwu, Modestus O., and Lagouge K. Tartibu. "Whale optimization algorithm (woa)." In Metaheuristic optimization: nature-inspired algorithms swarm and computational intelligence, theory and applications, Cham: Springer International Publishing, 2020, 53-60.