ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
Received: 20 March 2025; Accepted: 27 July 2025

https://doi.org/10.17559/TV-20250320002495
Original scientific paper

Kernel Extreme Learning Machine-Based Sentiment Analysis for Social Networks

Janani TAMILARASU*, Logeswari SHANMUGAM

Abstract: The exponential growth of user-generated content on social media platforms presents both opportunities and challenges in extracting meaningful insights. Sentiment Analysis (SA), a critical component of contextual mining, enables the identification of subjective information embedded within textual data. This article proposes a novel Kernel Extreme Learning Machine-Based Sentiment Analysis of Social Networks Using Improved Dung Beetle Optimization (KELMSASN-IDBO) model, which combines advanced machine learning and nature-inspired optimization techniques to enhance sentiment classification accuracy. The model follows a structured pipeline: initially, raw textual data undergo thorough preprocessing to eliminate noise and standardize content. Subsequently, semantic features are extracted using Bidirectional Encoder Representations from Transformers (BERT) for effective word embedding. The resulting features are then classified using a Kernel Extreme Learning Machine (KELM), known for its high generalization performance and rapid learning speed. To optimize the performance of KELM, an Improved Dung Beetle Optimization (IDBO) algorithm is employed for fine-tuning hyperparameters. Experimental results demonstrate that the proposed KELMSASN-IDBO model outperforms conventional sentiment analysis techniques in terms of accuracy, efficiency, and robustness. The integration of deep contextual embeddings and hybrid optimization makes the proposed model a powerful tool for extracting sentiments from complex and large-scale social network data.

Keywords: improved dung beetle optimization (IDBO); kernel extreme learning machine (KELM); natural language processing (NLP); sentiment analysis; social network analysis

1 INTRODUCTION

The development of the information technology has greatly changed the status of social media, including Facebook, Instagram, and Twitter, that are now an essential part of the contemporary lifestyle [1]. Such platforms have grown tremendously over the past years and have significantly affected the daily activities of people. Many users are active in these platforms not only to socialize and share personal experiences but also to share sentiments and opinions about different services, organizations and products in terms of posts and comments [2]. Therefore, a massive amount of user-generated information is constantly generated and has great importance to organizations, governments and individuals. It is important to derive and apply valuable insights out of this large data to make decisions and analyze trends [3]. Nevertheless, it is a difficult task to conduct a successful data analysis, and the automated sentiment and opinion extraction methods should be developed [4].

Sentiment Analysis (SA) has become a computational method that is aimed at extracting emotions, sentiments, and opinions automatically out of a textual data. The first benefit of SA is that it has the capability to categorize and analyze consumer attitudes, which can give useful information about how people feel about certain subject, service or product [5]. SA is the process of locating and classifying any opinions represented in a text to establish the position of the writer on a topic and it is highly linked to text mining [6]. SA and text analysis are frequently used interchangeably because they both entail the extraction of meaningful information in textual data. SA is mainly concerned with identifying and classifying sentiments in text-based inputs, usually by rating them according to polarity (e.g. positive, negative or neutral) [7].

NLP (Natural Language Processing) is a critical element of the SA, which allows computers to communicate with human language using intelligent models. Some linguistic analysis enabled by NLP is lexical analysis (establishing the structure of words in a sentence) and syntactic analysis (establishing grammar and word relationships) [8]. Also, the dictionary-based method in SA assists in drawing out the exact meaning of a piece of writing. Machine Learning (ML) models have gained more and more importance to SA in social networks where artificial intelligence (AI) is used to classify sentiments based on unsupervised, semi-supervised, supervised, and hybrid learning methods [9]. Notable supervised learning techniques of SA comprise Maximum Entropy Model, Naive Bayes (NB), and Support Vector Machines (SVM) [10].

This paper presents a Kernel Extreme Learning Machine-Based Sentiment analysis of Social Networks using Improved Dung Beetle Optimization (KELMSASN-IDBO) model, which aims at improving the accuracy and efficiency of SA. The offered framework contains a multi-stage text preprocessing pipeline that cleans raw text data to analyze it successfully. Word embedding is done using Bidirectional Encoder Representations of Transformers (BERT) to extract contextual meaning. The KELM model is used to classify data and the optimization of hyperparameters is achieved using the Improved Dung Beetle Optimization (IDBO) algorithm. The performance of the KELMSASN-IDBO model is validated through extensive simulations and this validates the effectiveness of the model in sentiment classification in social networks.

2 RELATED WORKS

The recent advancements in sentiment analysis (SA) have applied machine learning (ML) and deep learning (DL) to extract valuable information out of the social media data. In another research, the opinion mining was proposed as an SA model to predict the number of confirmed cases of infectious diseases by applying ML models. Word2Vec was applied to generate a sentiment dictionary and extend the existing one, and sentiment polarity was labeled as either positive or negative. Then, a predictive model was created to forecast the number of confirmed cases on the basis of Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU) and Deep Neural

Networks (DNN) that would utilize the sentiment trends [11].

Another approach introduced a Deep Learning-based Grasshopper Optimizer Algorithm-based sarcasm detection framework (SD-GOADL). This methodology was meant to determine the trends of sarcasm in the social media data using the Glove-based word embeddings and pre-processing methods. The classification was finished with the help of Deep Belief Network (DBN) and the Grasshopper Optimization Algorithm (GOA) was employed to optimize the hyperparameters to enhance the accuracy of detection [12].

Additional study created a sentiment analysis model on a low-resource document-level sentiment dataset gathered on social media. Rather than using the conventional handcrafted features, three main features were extracted with the help of word embeddings, and then the MC2D-CNN+BiLSTM-Attn, a new hybrid model, was used to classify the sentiment accurately [13].

The Support Vector Machines (SVM) and Naive Bayes (NB) machine learning methods have also been compared based on their ability to enhance the accuracy of SA. Performance was evaluated using a strict pre-processing pipeline, feature extraction and text normalization. Nevertheless, the results showed that there is a significant difference between the accuracy of various models, which implies that their efficiency is strongly conditioned by particular working conditions [14].

In another study, an unsupervised SA method of labeling Arabic text was proposed, where K-means clustering was used to identify the sentiment polarity on a multi-domain dataset of positive and negative reviews. The model incorporated pre-processing, feature selection, dimensionality reduction and clustering in order to enhance the classification performance [15].

The implementation of SA of multi-modal data has been proposed to be in the form of Multi-Attentional Fusion (MAF) system, which uses residual units and cross-attention. Textual and auxiliary modalities were combined during the fusion process that consisted of three pairs of twofold attention computations. A residual unit was also used to aggregate the information into attention module [16].

Further, a hybrid Arithmetic Optimizer Algorithm-Hunger Games Search (AOA-HGS) optimized Ensemble Multi-scale Residual Attention Network (EMRA-Net) model was developed to carry out multi-modal sentiment analysis. The AOA-HGS model could capture extensive and complementary features in most modalities, including text, social connections, video, and audio. The system was operating in two phases: TRA-CNN and EA-CNN, which enhanced the process of sentiment classification [17].

3 MATERIALS AND METHODS

It has been proposed that SA of multi-modal data be performed with the help of a Multi-Attentional Fusion (MAF) system that involves residual units and cross-attention. The fusion of the textual and auxiliary modalities was accomplished in the process of fusion that involved three pairs of twofold attention calculations. The attention module also aggregated the information via a residual unit [18].

Moreover, an Ensemble Multi-scale Residual Attention Network (EMRA-Net) model that was optimized using a hybrid Arithmetic Optimizer Algorithm-Hunger Games Search (AOA-HGS) was designed to conduct multi-modal sentiment analysis. The AOA-HGS model could represent holistic and complementary features in most modalities, including text, social relationships, video, and audio. The system was operating in two phases: TRA-CNN and EA-CNN that enhanced the sentiment classification procedure [19].

3.1 Text Pre-Processing

At first, the text pre-processing stage contains various levels to clean and transform raw text data into a suitable format that can be effectively employed for analysis. When texts are examined, numerous words do not have an influence on their sentiment alignment [20]. For instance, question words such as when, what, how do not offer an impact to the text polarity, hence they are eliminated to decrease the problem dimensionality as all words in the text are preserved as single dimension. In addition, particularly after texts originated from online SNs, they are in raw design and typically comprise too noisy, and inconsistent or incomplete data portions. Then a data pre-processing stage becomes essential, namely the process completed for cleaning and preparing the text for classification.

Despite significant advancements in sentiment analysis, several challenges remain. Existing models rely on traditional word embeddings such as Word2Vec and Glove, which struggle to capture deep semantic relationships in social media text . Deep learning models like LSTM and GRU exhibit high computational costs, limiting real-time processing. Additionally, current hyperparameter tuning techniques, such as GOA and K-means, lack efficiency in optimizing classification models [21]. Most approaches fail to integrate multi-modal social media features, reducing generalization. To overcome these issues, this study proposes a Kernel Extreme Learning Machine-Based Sentiment Analysis Model with Improved Dung Beetle Optimization for enhanced accuracy and efficiency.

Pre-processing in SA is akin to conventional text pre-processing in text mining. General pre-processing stages in text mining contain:

replacing or removing stemming: A model which lessens words to their general stem, or root;

lemmatization: A stemming-related method groups together the dissimilar changed formats of the word, such as walk, walking, and walked such that they are examined as a particular item;

Tokenization: The splitting of the stream of text into small components termed tokens;

Stop words removal: The method which eliminates words such as determinants like the a, an, other, directing combinations as yet, nor, so, for, an, or, but and prepositions namely under, in, before, toward;

Negation handling.

When input data originates from SNs, pre-processing needs numerous stages, like online text cleaning (such as eliminating URLs, the Retweets tag [RT] or HTML tags), increasing acronyms or abbreviation, removing or handling emoticons, and removing or replacing recurrent characters such as the o, p and y in the succeeding words, cooooooool or happyyyyyyyyyyyyyyyyy.

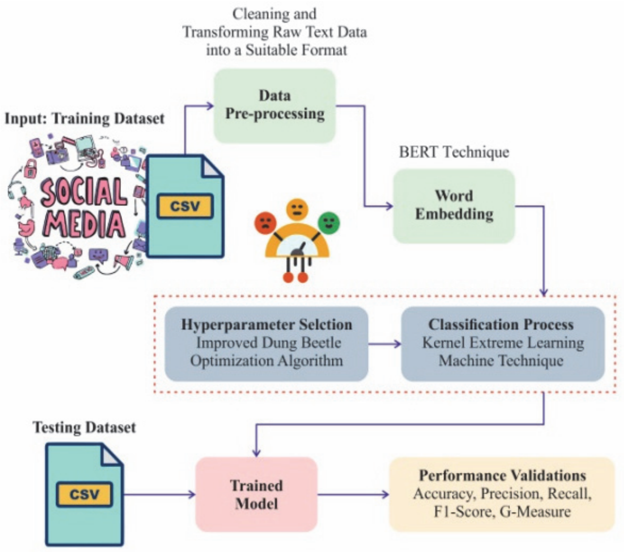

Figure 1 Overall working process of KELMSASN-IDBO model

The preprocessing stage in the present research takes a multi-level approach to converting raw social media text into a clean and analyzable form. First, the usual NLP preprocessing is used, consisting of tokenization (dividing text into words or phrases), lemmatization (reducing words to their base form), and stemming (cutting words to root forms). Stop words like is, the and and are eliminated since they are not used to determine sentiment polarity. In addition, special attention is paid to cleaning of social media specifics, including the deletion of URLs, usernames, hashtags, emojis, HTML tags, and retweet symbols. Acronyms and abbreviations (e.g. lol, brb) are expanded based on a lookup table and elongated words (e.g. soooo) are normalized. Sentiment context is also maintained by negation handling. The given holistic preprocessing pipeline guarantees that the text input is noise-free and properly structured to undergo further BERT-based contextual embedding.

3.2 Word Embedding

The choice of BERT to perform word embedding is based on the bidirectional contextual comprehension that enables the model to capture the semantic nuances more efficiently than the traditional embedding techniques such as Word2Vec or GloVe. In contrast to these fixed embeddings, BERT is contextual and dynamically creates representations of words depending upon the context of a word in the sentence, which is especially useful in sentiment analysis where the meaning of a word can vary significantly depending upon its usage. Moreover, the WordPiece tokenization used in BERT is graceful to out-of-vocabulary terms, which are prevalent in social media texts.

Besides, the proposed KELMSASN-IDBO model executes the BERT technique for word embedding process. BERT is a pre-trained language method made to improve the efficiency and quality of NLP solutions. In addition, it is successfully applied in numerous NLP tasks, comprising text classification and question answering [19]. It is designed to understand the word's contextual meaning inside a word by examining it bi-directionally - from right to left and left to right. This attribute reduces BERT remarkably intelligent for the depiction of text tasks, like sentiment analysis (SA), whereas understanding refined connections among words is important. BERT uses WordPiece Tokenization which analyses text into small sub-word components for managing out-of-vocabulary or rare terms. For sample, the term unhappiness might be separated into un, ##happy, and ##ness. This sub-word tokenization permits BERT to precisely characterize strange words whereas preserving context importance. Specific tokens in BERT [CLS] Characterize the whole order and is applied for sentence-level jobs (for example: sentiment categorization). [SEP] differentiates phrases or sentences in jobs like question replying or following sentence estimation. For example, the expression The product is amazing! is tokenized as: [CLS] The product is amazing! [SEP]

All tokens are characterized as an arrangement of the following embedding's:

Token embedding: word embedding's characterized as numeric vectors.

Segment embedding: The fixed model, which defines whether the token is portion of the 1st or 2nd sentences after the final is given as input.

Position embedding: The position representations for all characters inside the sentence. Regarding dual sentences, the location in the next sentences emulates successively from the last location of the 1st sentence, raised by a level.

BERT is constructed on the transformer structure, containing the next core elements, Initial Encoding Layers, it utilizes numerous transformer encoders (for example: twelve inBERT_BASE), All layers handle the sequence of input to take word relations. Formerly Multi Head Attention Mechanism, takes the significance of all words within the sequence in relation to another word, permitting the method to concentrate on important portions of the input. Then Feed-Forward NNs improve the context description of words. Lastly Hidden Layers (HLs), All encoding layers outputting a 768-dimension vector for all tokens (inBERT_BASE).

For SA, the [CLS] token's last HL is frequently applied as feature representations for the complete sequence of input. This vector takes the meaning and context of the input text. The initial method is Tokenization of the text into Piece of Word tokens then Passing the tokenized input over BERT and remove the [CLS] output of the token being the feature vector for the input text. Lastly utilize this vector as input to a down-stream classifier to forecast sentiment.

3.3 Classification Process

Sentiment classification in the proposed KELMSASN-IDBO model is done with the help of the Kernel Extreme Learning Machine (KELM) which is a variation of the standard Extreme Learning Machine (ELM). The ELM is a feedforward neural network which has only one layer of hidden nodes. Among its main strengths, it has a rapid training algorithm that does not use iterative weight updates that are typical of other neural networks. Rather, the weights between the hidden and input layers are randomly selected and held constant, and the weights between the output and the hidden layers are calculated analytically to provide the best fit to the training data.

Standard ELM can however be unstable and perform inconsistently because of the random initialization of parameters. To address this, the KELM is an extension of ELM that incorporates kernel methods to enable the model to map input data to a high dimensional feature space in which complex data patterns can be linearly separated. This avoids the necessity of explicitly selecting the weights and biases of the hidden layers, and the training becomes more stable and the model more able to deal with nonlinear relationships in the data.The kernel function of KELM calculates the similarity of data instances which allows the model to learn complex, non-linearly separable sentiment patterns using small numbers of neurons and layers. Some typical kernel functions are radial basis function (RBF), polynomial and linear kernels. These functions are useful in encoding the inherent organization of the input data, which can contain sentiment clues that can be implicit or contextual-particularly in the case of social media text, which can contain sarcasm or irony or ambiguous language.

KELM uses the labeled sentiment data in a training process to identify an optimal mapping of the feature representations that are generated by the BERT embedding to the sentiment classes that they belong to. When the model is trained it can be used to make predictions on unseen, new text data using this learned mapping. KELM is particularly suitable in the real-time sentiment analysis due to its fast training and high generalization capacity. To further enhance the classification performance, the hyper parameters of the KELM model such as the kind of kernel and the regularization parameters are optimized using the aid of the Improved Dung Beetle Optimization (IDBO) algorithm. This is optimized such that the classifier is executed using parameters that provide the highest accuracy of prediction on the validation set.

3.4 Hypermeter Tuning Model

Hyperparameter tuning is a significant part of any machine learning model that determines the performance and the generalization ability of the classifier. It comprises finding the optimal set of model parameters that are not discovered by the data as such, but rather are chosen in advance of the training procedure. In the proposed KELMSASN-IDBO framework using the Kernel Extreme Learning Machine (KELM) classifier, the type of the kernel, kernel parameters (e.g., spread or degree) and regularization coefficients are hyperparameters and highly influence the accuracy of the classification, robustness and convergence behavior. Improper tuning of these values can lead to overfitting, underfitting or slow convergence issue, which will decrease the performance of the model.

To eliminate these challenges, the paper applies an effective metaheuristic optimization algorithm, the Improved Dung Beetle Optimization (IDBO) algorithm. This is a bio-inspired optimization algorithm which is inspired by the behavior of the dung beetles which are intelligent in movement and decision making in identifying the best paths and resources. This improved variant has several strategic modifications that improve search diversity, local optima avoidance, and convergence speed, which makes it especially suitable to use in the case of complex hyperparameter optimization tasks in sentiment classification tasks.

3.4.1 Rationale for Using Metaheuristic Optimization

The more conventional hyperparameter optimization approaches such as grid search and random search are computationally expensive and often ineffective in high-dimensional or continuous hyperparameter spaces. These methods may either sweep systematically through grids of fixed parameters, or sample randomly combinations of parameters, and both may fail to find the optimal settings due to insufficient sampling or due to the absence of intelligent guidance.

Metaheuristic optimization algorithms, in contrast, are built to intelligently explore and exploit the search space by emulating biological evolution, natural phenomena or swarm behaviour. These are stochastic and population based algorithms that provide a balance between global exploration (searching new areas) and local exploitation (polishing known good solutions). They are especially effective in non-convex, multi-modal, and noisy optimization landscapes that are typical in machine learning tasks in the real world.

The Dung Beetle Optimization (DBO) algorithm is one of the metaheuristics that have demonstrated potential because of its easy implementation, low parameter tuning constraints, and high convergence qualities. The enhanced version (IDBO) that has been used in the present study is a modification of the original concepts of DBO and incorporates several enhancements to overcome some of the limitations that are inherent to DBO.

3.4.2 Behavior-Inspired Improvements in IDBO

The Improved Dung Beetle Optimization algorithm brings in some important mechanisms to improve performance:

1. Refractive Reverse Learning (RRL):

Premature convergence is one of the most important problems in standard optimization methods, i.e. the algorithm converges to a local optimum. This is overcome by the RRL mechanism which produces oppositely (or refracted) alternative candidate solutions to the current best solution. This will widen the area of search and will enhance the likelihood of finding superior solutions in other areas of the search space that have not been explored.

2. Adaptive Population Balancing:

Dung beetles in the natural environments have various functions including small dung beetles (which forage widely) and breeding dung beetles (which utilize local resources). Based on this, the IDBO algorithm has proposed an adaptive control mechanism in which the proportion of exploration and exploitation agents are changed dynamically as the iteration progresses. During the early phase, there are more agents tasked with exploration in order to find good areas. Convergence leads to exploitation as convergence proceeds with the aim of perfecting the best solutions. This dynamic balance enhances efficiency and reliability of the search.

3. Subtractive Averaging Strategy:

To enhance further search capability globally, a subtractive averaging mechanism is added in the rolling process by the algorithm. In this method, the average difference between the members of the population is calculated and the positions are adjusted in this way, the agents are motivated to move to more promising areas according to the collective behavior, instead of individual fitness. It improves collaborative learning behavior of the population, eliminates redundancy, and increases speed of convergence.

4. Obstacle-Aware Direction Adjustment:

Based on the behavior of the beetles, the algorithm simulates scenarios when beetles face an obstacle and engage in a dancing behavior to shift the direction. This is replicated in IDBO by randomly perturbing and re-initialization strategies when stagnation or performance loss is observed. This also assists the optimizer to get out of local optima and keep the population diverse.

3.4.3 IDBO for KELM Hyperparameter Tuning

In the KELMSASN-IDBO framework, the IDBO algorithm is applied to optimize a number of hyperparameters of the Kernel Extreme Learning Machine such as:

- Type of the kernel (e.g. RBF, polynomial, sigmoid),
- Parameters that are kernel-specific (e.g. the width of the RBF kernel or the degree of the polynomial kernel),
- Regularization coefficient, that regulates the trade-off between training error minimization and preservation of generalization capability.

A given set of these hyperparameters is represented by each candidate solution in the IDBO population. A fitness function is used to measure the quality of each candidate, and in this case is determined by the classification performance (e.g. precision or F1-score) of the KELM classifier on a validation subset of the training data. This is to maximize this fitness so that the hyperparameters chosen lead to the most correct and robust sentiment classification.

The search starts at a random set of hyperparameter combinations. During every iteration, the IDBO algorithm modifies the population according to its improved behavioral strategies (as explained above), generates new solutions, and selects the best individuals. It is repeated a certain number of times or until convergence (e.g. no improvement between successive generations).

3.4.4 Integration with the Overall Sentiment Analysis Pipeline

The KELMSASN-IDBO model hyperparameter tuning procedure is closely connected with the end-to-end sentiment analysis pipeline. After the BERT-based word embedding layer, the fine-tuned KELM classifier, which is tuned through IDBO, is utilized to conduct sentiment classification. This smooth integration makes sure that the advantages of each part (contextual embedding, non-linear classification, and metaheuristic optimization) are used in a combined way, which leads to a high-performance and scalable sentiment analysis solution. Also, the IDBO is a wrapper around the classifier, so it does not introduce any limitations on the format of the data or type of the task. This renders it flexible to a variety of sentiment analysis issues, such as binary, multi-class, or even regression-based sentiment forecasting.

4 PERFORMANCE ANALYSIS

Here, the performance evaluation of the KELMSASN-IDBO technique is examined under the emotions dataset [22]. The dataset contains 60000 samples under six emotions as exposed in Tab. 1. Tab. 2 provides sample texts of emotions. The emotion classes are Sadness, Joy, Love, Anger, Fear, and Surprise. Every class will contain 10,000 text samples. The data set is balanced and diverse in terms of the emotional content, which makes it a good choice when it comes to multi-class sentiment classification. All the samples are preprocessed with the full pipeline outlined in Section 3.1 before being embedded with BERT, to facilitate high-quality inputs to robust performance evaluation.

Table 1 Details of dataset

Emotions	Samples
Sadness (0)	10000
Joy (1)	10000
Love (2)	10000
Anger (3)	10000
Fear (4)	10000
Surprise (5)	10000
Total	60000

Table 2 Sample texts of emotions

S.no	Emotions	Text
1	Sadness (0)	realizing that school will soon be over
2	Joy (1)	*i* feel valued when i_m with him
3	Love (2)	*i* watch the film *i* feel sympathetic for all the characters
4	Anger (3)	*i* knew that feeling and *i* felt disgusted at myself for feeling it
5	Fear (4)	*i* just feel really helpless and heavy hearted
6	Surprise (5)	*i* feel like that combo is kinda weird

Fig. 3 shows an overall analysis of the classification performance of the KELMSASN-IDBO model in terms of four subplots. Subfigs. 3a and 3b represent the confusion matrices of the training (70% TRPH) and testing (30% TSPH) data, respectively. These matrices indicate that the model has a high class-wise accuracy and most of the instances are correctly classified in all the six emotion classes. It is important to note that the diagonal dominance of the matrices is rather high, which proves the effectiveness of the model to differentiate subtle differences in emotional sentiments, like Joy vs. Love or Anger vs. Fear. Fig. 3c demonstrates the Precision-Recall (PR) curve where the precision and recall are high and consistent in all the emotion categories. The curves are high above baseline and this fact proves that the model has a good balance between the identification of true positives and the avoidance of false ones. The Receiver Operating Characteristic (ROC) curves of the multiclass classification are presented in Fig. 3d. The true positive and false positive rates of all class curves are high and low respectively, resulting in Area Under Curve (AUC) values that are near 1.0. This shows that the classifier is discriminative and strong to different tones of emotions even in the smooth transitions of sentiments.

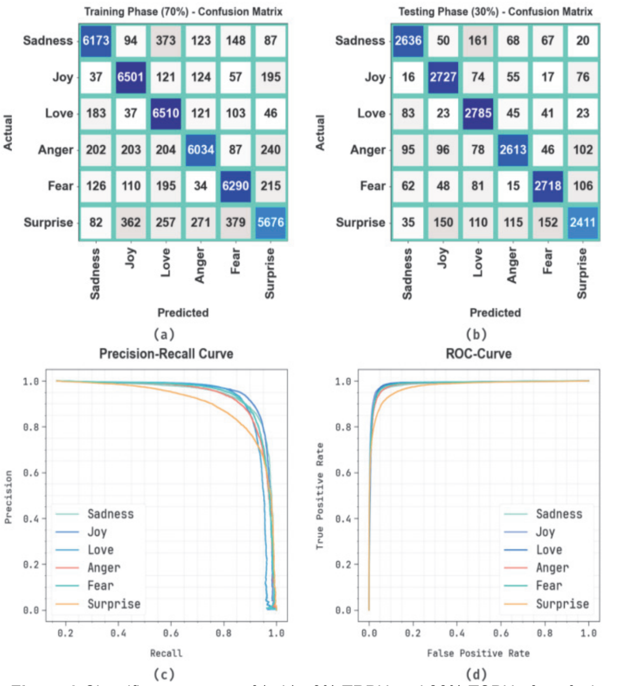

Figure 3 Classifier outcomes of (a-b) 70% TRPH and 30% TSPH of confusion matrix and (c-d) curves of PR and ROC

Tab. 3 provides a class-wise performance measure in detail. The model records especially high accuracy in all the categories of emotions, with Joy (96.81%) and Fear (96.54%) being some of the highest-performing classes. Interestingly, the values of precision and recall are slightly lower in Surprise, which is probably caused by the ambiguous or context-dependent nature of the textual data. However, the F1-scores are quite strong in all categories, showing equal performance in both recognizing correctly and reducing errors per emotion category. The standard deviation between metrics is also low, which once again proves the stability of the model.

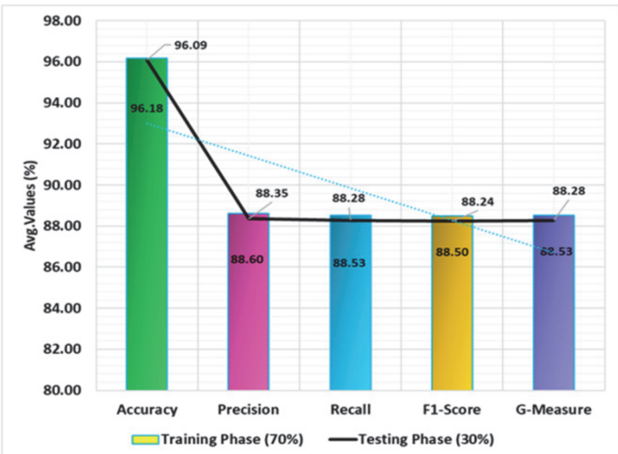

Figure 4 Average of KELMSASN-IDBO model under 70% TRPH and 30% TSPH

Fig. 4 illustrates the mean classification measures of KELMSASN-IDBO model using two data divisions 70 percent training (TRPH) and 30 percent testing (TSPH). The metrics are Accuracy, Precision, Recall, Specificity and F1-score. In both splits, all of the metrics are above 88%, with Accuracy being close to 96%, which indicates that the model generalizes well and is able to perform highly on unseen data. The fact that training and testing scores are very close also means that the model does not overfit, which is the advantage of both BERT as a contextual representation and IDBO as an efficient hyperparameter tuning.

Table 3 Emotions detection of KELMSASN-IDBO model under 70% TRPH and 30% TSPH

Class Labels	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{score}$	$G_{measure}$
TRPH (70%)					
Sadness	96.54	90.74	88.21	89.46	89.47
Joy	96.81	88.97	92.41	90.66	90.67
Love	96.10	84.99	93.00	88.81	88.90
Anger	96.17	89.97	86.57	88.24	88.25
Fear	96.54	89.04	90.24	89.64	89.64
Surprise	94.92	87.88	80.77	84.18	84.25
Average	96.18	88.60	88.53	88.50	88.53
TSPH (30%)					
Sadness	96.35	90.06	87.81	88.92	88.93
Joy	96.64	88.14	91.97	90.01	90.04
Love	96.01	84.68	92.83	88.57	88.66
Anger	96.03	89.76	86.24	87.96	87.98
Fear	96.47	89.38	89.70	89.54	89.54
Surprise	95.06	88.06	81.10	84.43	84.51
Average	96.09	88.35	88.28	88.24	88.28

Besides the overall classification accuracy, the evaluation will provide a full range of performance measures: Precision, Recall, F1-score, and Specificity per each of the six emotion classes. These measures give a more detailed picture of how well the model performs in identifying genuine positive sentiments whilst reducing false positive and false negative sentiments. The KELMSASN-IDBO model is very precise and recalls in both training (70%) and testing (30%) splits, as indicated in Tab. 3 and Fig. 4, with F1-scores greater than 88% in the majority of emotion classes. The presence of these detailed metrics guarantees that the performance of the classifier will be measured not only on the correctness but also on the robustness of different sentiment categories.

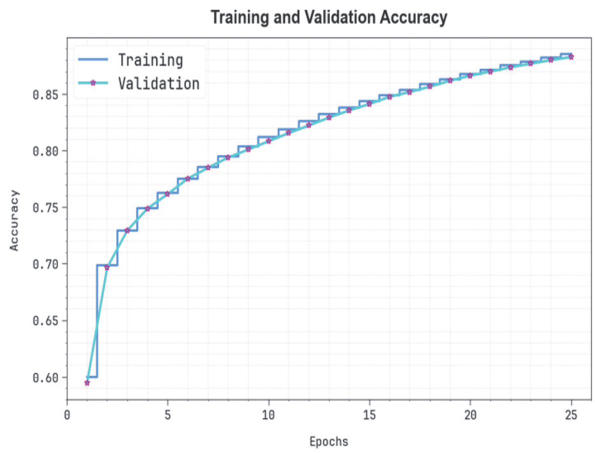

Figure 5 $Accu_y$ curve of KELMSASN-IDBO model

The trend of training and validation accuracy is shown in Fig. 5 over 25 epochs. The two curves have an upward consistent trend, and the validation accuracy follows the training accuracy closely across the epochs. This action shows the ability of the model to learn well and have stable generalization. The fact that both curves overlap also indicates the lack of overfitting, which proves the BERT embeddings and KELM classifier are aligning well with the IDBO-optimized parameter set.

Another important strength of the proposed method is its computational efficiency of the KELMSASN-IDBO model. Kernel Extreme Learning Machine (KELM) is computationally cheaper as it does not involve the iterative tuning of weights, training is faster than deep neural networks. Its main computational cost is the computation of the Moore-Penrose pseudo-inverse, and hence its complexity is $O(N3)$, where N is the number of samples. This is however offset by its single pass training and generalization. Although the IDBO algorithm introduces overhead in the form of population-based search, it works with a complexity of $O(P, D, I)$, where P is the number of agents, D is the dimension of the solution, and I is the number of iterations. The enhanced convergence of IDBO guarantees that the optimal hyperparameters can be identified using a limited number of iterations rather than the traditional optimization methods. In general, the model is highly accurate with reasonable training times, which makes it suitable to real-time sentiment analysis applications.

Fig. 6 shows the value of loss during training and validation phases over 25 epochs. The downward sloping nature of the two curves indicates that there is constant learning and a gradual decrease in error. In addition, the low difference between the training and validation loss reflects good generalization and little variance between the datasets.

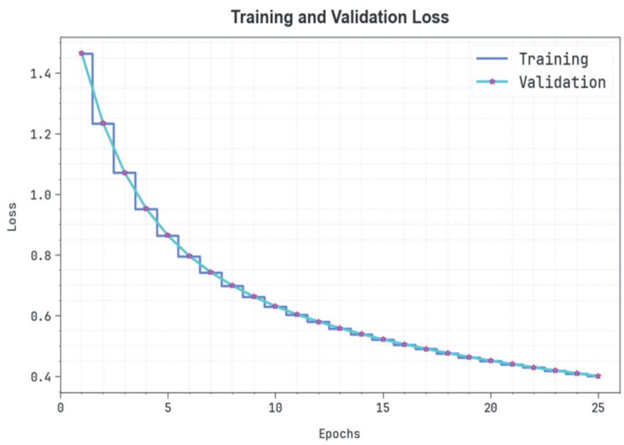

Figure 6 Loss curve of KELMSASN-IDBO model

Tab. 4 shows a direct comparison of KELMSASN-IDBO with a number of well-established classifiers, namely, SVM, GNB, ETC, KNN, 1D-CNN, MLP, and Wav2Vec2-XLSR [23, 24]. The suggested model performs much better than any baseline in the main metrics:

Precision: KELMSASN-IDBO has a 96.18% score, which is higher than that of even the nearest rival, Wav2Vec2-XLSR (95.50%).

Precision and Recall: The two measures are above 88%, in contrast to others, including GNB (81-85%) and KNN (81-86%).

F1-Score: KELMSASN-IDBO provides 88.50%, which is a high trade-off between precision and recall compared to the second best (Wav2Vec2-XLSR at 80.12%).

Fig. 7 gives a visual representation of the performance measures of all models in Tab. 4. The bars corresponding to KELMSASN-IDBO are always on the top of all categories, which is an indicator of its high classification ability. This graphical representation confirms the strength and flexibility of this model. It also shows that lightweight architectures such as KELM can achieve better accuracy and efficiency compared to deep ones when they are optimized.

Table 4 Comparative analysis of KELMSASN-IDBO technique with existing approaches

Model	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{score}$
SVM Classifier	90.06	86.07	87.05	87.07
GNB Model	88.06	81.06	84.08	85.05
ETC Method	85.07	87.07	77.06	81.07
KNN Algorithm	89.06	81.07	86.06	83.08
1D-CNN	92.52	81.24	82.14	79.73
MLP Model	93.99	82.22	85.31	81.52
Wav2Vec2-XLSR	95.50	85.07	84.99	80.12
KELMSASN-IDBO	96.18	88.60	88.53	88.50

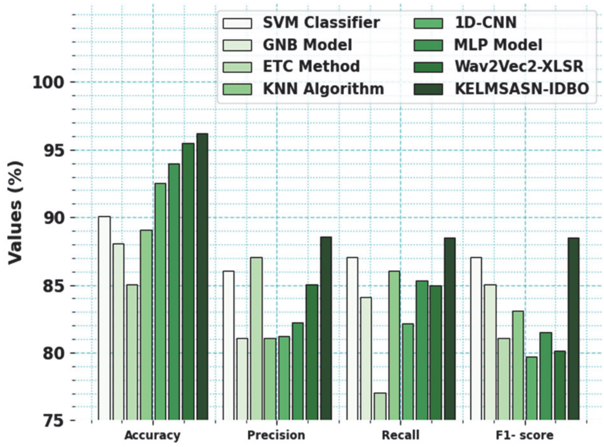

Figure 7 Comparative analysis of KELMSASN-IDBO method with existing approaches

5 CONCLUSION

The Kernel Extreme Learning Machine-Based Sentiment Analysis of Social Networks Using Improved Dung Beetle Optimization (KELMSASN-IDBO) model introduced in this paper combines the state-of-art machine learning approaches and optimization algorithms to improve sentiment analysis of social networks. The model starts with a successful text pre-processing step, which guarantees the conversion of raw text information to a structured form that can be analyzed. The Bidirectional Encoder Representations from Transformers (BERT) method is used to provide word embedding to capture deep contextual representations of text, which increases the quality of feature extraction significantly. In the classification of sentiment, the Kernel Extreme Learning Machine (KELM) method is applied because it offers high learning speed and excellent generalization capability. But the performance of KELM is highly dependent on the best choice of hyperparameters which are solved by Improved Dung Beetle Optimization (IDBO) algorithm. The IDBO algorithm optimizes the hyperparameters, and thus improves the classification performance, resulting in more accurate sentiments. Wide experimental analysis shows that the KELMSASN-IDBO model is more effective than the current sentiment analysis methods. The findings emphasize its capability to attain a greater level of classification accuracy, lesser complexity of computation, and enhanced robustness in the analysis of social media sentiments. The combination of BERT-based word embedding, KELM classification, and IDBO-based optimization guarantees an efficient sentiment analysis framework and its scalability. On the whole, the suggested KELMSASN-IDBO model can be regarded as a hopeful solution to the automated sentiment classification in social networks, which is why it can be used in market analysis, customer feedback analysis, and monitoring of public opinion. In future, the study can extend this model to multi-modal sentiment analysis by including more features like image and video data in order to make it more applicable in real life. Future improvements will be the extension of the model to multi-modal sentiment analysis, where features of images, videos, and audio contents are added.

6 REFERENCES

[1] Baydogan, C. & Alatas, B. (2021). Sentiment Analysis in Social Networks Using Social Spider Optimization Algorithm. *Tehnički vjesnik/Technical Gazette*, *28*(6). https://doi.org/10.17559/TV-20200614172445

[2] Yue, L., Chen, W., Li, X., Zuo, W., & Yin, M. (2019). A survey of sentiment analysis in social media. *Knowledge and Information Systems*, *60*(2), 617-663. https://doi.org/10.1007/s10115-018-1236-4

[3] Chauhan, P., Sharma, N., & Sikka, G. (2021). The emergence of social media data and sentiment analysis in election prediction. *Journal of Ambient Intelligence and Humanized Computing*, *12*(3), 2601-2627. https://doi.org/10.1007/s12652-020-02423-y

[4] Nemes, L. & Kiss, A. (2021). Social media sentiment analysis based on COVID-19. *Journal of Information and Telecommunication*, *5*(1), 1-15. https://doi.org/10.1080/24751839.2020.1790793

[5] Rajagopal, S., Jayaprakash, S., & Subburathinam, K. (2024). Sentiment Analysis on Big Data: A Hybrid SED-TABU Feature Selection Method. *Tehnički vjesnik/Technical Gazette*, *31*(6), 2079-2086. https://doi.org/10.17559/TV-20231004000989

[6] Drus, Z. & Khalid, H. (2019). Sentiment analysis in social media and its application: Systematic literature review. *Procedia Computer Science*, *161*, 707-714. https://doi.org/10.1016/j.procs.2019.11.174

[7] Etter, M., Colleoni, E., Illia, L., Meggiorin, K., & D'Eugenio, A. (2018). Measuring organizational legitimacy in social media: Assessing citizens' judgments with sentiment analysis. *Business & Society*, *57*(1), 60-97. https://doi.org/10.1177/0007650316683926

[8] Shayaa, S., Jaafar, N. I., Bahri, S., Sulaiman, A., Wai, P. S., Chung, Y. W., Piprani, A. Z., & Al-Garadi, M. A. (2018). Sentiment analysis of big data: methods, applications, and open challenges. *IEEE Access*, *6*, 37807-37827. https://doi.org/10.1109/ACCESS.2018.2851311

[9] Chakraborty, K., Bhattacharyya, S., & Bag, R. (2020). A survey of sentiment analysis from social media data. *IEEE Transactions on Computational Social Systems*, *7*(2), 450-464. https://doi.org/10.1109/TCSS.2019.2956957

[10] Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R., & Mora, H. (2020). A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. *Industrial Marketing Management*, *90*, 523-537. https://doi.org/10.1016/j.indmarman.2019.08.003

[11] Melton, C. A., Olusanya, O. A., Ammar, N., & Shaban-Nejad, A. (2021). Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence. *Journal of Infection and Public Health*, *14*(10), 1505-1512. https://doi.org/10.1016/j.jiph.2021.08.010

[12] Song, Y. & Yoon, B. (2024). Prediction of infectious diseases using sentiment analysis on social media data. *Plos one*, *19*(9), e0309842. https://doi.org/10.1371/journal.pone.0309842

[13] Srinu, N., Sivaraman, K., & Sriram, M. (2024). Enhancing sarcasm detection through grasshopper optimization with deep learning based sentiment analysis on social media. *International Journal of Information Technology*, 1-7. https://doi.org/10.1007/s41870-024-02057-9

[14] Agbesi, V. K., Chen, W., Ukwuoma, C. C., Kuadey, N. A., Agbesi, C. C. M., Ejiyi, C. J., Gyarteng, E. S., Muoka, G. W., & Kuadey, A. M. (2025). Multichannel 2D-CNN attention-based BiLSTM method for low-resource ewe sentiment analysis. *Journal of Data Science and Intelligent Systems*, *3*(1), 67-77. https://doi.org/10.47852/bonviewJDSIS32021512

[15] Yogi, K. S., Gowda, V. D., Sindhu, D., Soni, H., Mukherjee, S., & Madhu, G. C. (2024, April). Enhancing Accuracy in Social Media Sentiment Analysis through Comparative Studies using Machine Learning Techniques. *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)*, *1*, 1-6. https://doi.org/10.1109/ICKECS61492.2024.10616441

[16] Huang, J., Tong, Y., Hang, B., Xu, D., Wang, F., & Yu, J. (2024). FBF: A High-Efficient Query Mechanism for Keyword Search in Online Social Networks. *Tehnički vjesnik/Technical Gazette*, *31*(2), 352-361. https://doi.org/10.17559/TV-20230511000622

[17] Liu, Z., Yang, T., Chen, W., Chen, J., Li, Q., & Zhang, J. (2024). Sentiment analysis of social media comments based on multimodal attention fusion network. *Applied Soft Computing*, *164*, 112011. https://doi.org/10.1016/j.asoc.2024.112011

[18] Subbaiah, B., Murugesan, K., Saravanan, P., & Marudhamuthu, K. (2024). An efficient multimodal sentiment analysis in social media using hybrid optimal multi-scale residual attention network. *Artificial Intelligence Review*, *57*(2), 34. https://doi.org/10.1007/s10462-023-10645-7

[19] Zucco, C., Calabrese, B., Agapito, G., Guzzi, P. H., & Cannataro, M. (2020). Sentiment analysis for mining texts and social networks data: Methods and tools. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(1), e1333. https://doi.org/10.1002/widm.1333

[20] Taktack, F. & Demir, H. (2021). Relations that Show the Network Potential for Spatial Data Sharing. *Tehnički vjesnik/Technical Gazette*, *28*(6), 1943-1951. https://doi.org/10.17559/TV-20170920100230

[21] Gao, H., Xu, T., & Zhang, N. (2025). Robust kernel extreme learning machines for postgraduate learning performance prediction. *Heliyon*, *11*(1). https://doi.org/10.1016/j.heliyon.2024.e40919

[22] https://www.kaggle.com/datasets/bhavikjikadara/emotions-dataset

Contact information:

Janani TAMILARASU
(Corresponding author)
Department of Information Technology, Bannari Amman Institute of Technology,
Tamilnadu, India
E-mail: tamilarasujanani@gmail.com

Logeswari SHANMUGAM
Department of Information Technology, Karpagam college of Engineering,
Tamilnadu, India
E-mail: slogesh76@gmail.com