Received 18 October 2023, accepted 9 November 2023, date of publication 20 November 2023, date of current version 1 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3334980

RESEARCH ARTICLE

Sentiment Analysis Based on Hybrid Neural Network Techniques Using Binary Coordinate Ascent Algorithm

MOHAMMED HUSSEIN ABDALLA[1], JAFAR MAJIDPOUR[1], RAHELA AZIZ RASUL[2], ABDULRAHMAN A. ALSEWARI[3], (Senior Member, IEEE), TARIK AHMED RASHID[4], (Member, IEEE), ARAM MAHMOOD AHMED[5,6], BRYAR A. HASSAN[5], NOOR BAHJAT TAYFOR[7], SHKO MUHAMMED QADER[8,9], AND SINAN Q. SALIH[10]

[1]Department of Computer Science, University of Raparin, Rania 46012, Iraq
[2]Department of Mathematics, University of Raparin, Rania 46012, Iraq
[3]College of Computing, Faculty of Computing, Engineering and the Built Environment, Birmingham City University, B5 5JU Birmingham, U.K.
[4]Department of Computer Science and Engineering, University of Kurdistan Hewlêr, Erbil 44001, Iraq
[5]Department of Computer Science, College of Science, Charmo University, Sulaimani 46023, Iraq
[6]Department of Information Technology, University of Human Development, Sulaimani 46001, Iraq
[7]Department of Information Technology, Kurdistan Technical Institute, Sulaimani 46001, Iraq
[8]Department of Information Technology, University College of Goizha, Sulaimani 46001, Iraq
[9]Information Technology Department, Computer Science Institute, Sulaimani Polytechnic University, Sulaimani 46002, Iraq
[10]Technical College of Engineering, Al-Bayan University, Baghdad 10011, Iraq

Corresponding author: Sinan Q. Salih (sinan_salih@outlook.com)

ABSTRACT Sentiment analysis is a technique for determining whether data is positive, negative, or neutral using Natural Language Processing (NLP). The particular challenge in classifying huge amounts of data is that it takes a long time and requires the employment of specialist human resources. Various deep learning techniques have been employed by different researchers to train and classify different datasets with varying outcomes. However, the results are not satisfactory. To address this challenge, this paper proposes a novel Sentiment Analysis approach based on Hybrid Neural Network Techniques. The preprocessing step is first applied to the Amazon Fine Food Reviews dataset in our architecture, which includes a number of data cleaning and text normalization techniques. The word embedding technique is then used to capture the semantics of the input by clustering semantically related inputs in the embedding space on the cleaned dataset. Finally, generated features were classified using three different deep learning techniques, including Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), and Hybrid CNN-RNN models, in two different ways, with each technique as follows: classification on the original feature set and classification on the reduced feature set based on Binary Coordinate Ascent (BCA) and Optimal Coordinate Ascent (OCA). The experimental results show that a hybrid CNN-RNN with the BCA and OCA algorithms outperforms state-of-the-art methods with 97.91% accuracy.

INDEX TERMS Sentiment analysis, deep learning, BCA algorithm, OCA, RNN, CNN.

I. INTRODUCTION

With the development of science and technology, the use of the Internet has become essential for those who want to buy products or services online. All businesses who wish to sell their products put them on their websites and allow customers to provide feedback on the items they bought. They are the most effective tools for gathering information on people's opinions and sentiments on various topics. If someone wants to buy a product but doesn't know enough about

The associate editor coordinating the review of this manuscript and approving it for publication was Angel F. García-Fernández.

the pricing, quality, or after-sales support, they can ask their friends or relatives for assistance, nonetheless, they may not be familiar with the product. Today, all of the necessary information about each product is available on websites that applicants and users can access, and they may provide the required product with more confidence by studying the product's details as well as the opinions of the users who gave that product.

Analyzing a lot of consumer input on a product is the main problem. Manual extraction and opinion analysis are impossible since the data is unstructured and given in natural language. Text classification is a method for automatically assessing views called sentiment analysis. Text categorization is an important NLP activity that may be used in a variety of applications to decode natural language and solve problems [1]. Machine learning can analyze Twitter sentiments. Deep learning models in computer vision and voice recognition have performed well in recent years [2], [3]. Machine learning helps with NLP difficulties since it employs a generic learning algorithm and a large data sample to establish categorization rules.

CNN, a machine learning model, has recently excelled in natural language processing, adding to its previous success in image identification. Words may be considered jointly in a convolutional layer [4]. A RNN [5] contains neuron connections that loop back on themselves in a preset manner. Unlike feedforward neural networks, RNNs may utilize their internal "memory" to handle sequential inputs. Since they perform the same operation on each sequence element, RNNs have "memory" [6]. We offer a strategy for obtaining actionable insights from Twitter data utilizing CNN, RNN, and hybrid CNN-RNN models for sentiment analysis.

Sentiment analysis uses written remarks to determine how people feel about a topic or product. Sentiment analysis has grown with Twitter, blogs, and opinion polls. Many academics have examined feature selection as another strategy. The optimal number of characteristics for learning-based sentiment analysis remains unsolved. In this study, we applied the BCA and OCA methods to extract the best features and benefits of these methods, listed as follow:

i. If the model is provided with the right set of variables, it will improve the accuracy.
ii. The performance of deep learning algorithms in the classifier can be faster.
iii. The model complexity is reduced and interpretation is also easy.
iv. Overfitting can be reduced.

Today, most data are unstructured. Text categorization techniques are needed to process and analyze social media comments, consumer feedback, and browser traces. Deep learning enables software programs to self-train to perform specific tasks to expose neural networks to massive datasets [7]. Social media has made ideation and sharing easier. Scientific and marketing organizations sought public comment on social events, marketing activities, and product priorities [8]. Today, one may purchase a product without relying on family and friends' judgments since websites provide customer reviews and product conversations. Due to the diversity of viewpoints, discovering and assessing them is difficult. Automated sentiment analysis techniques are needed because sites include numerous text comments that are hard to decipher [9]. Social media's ability to convey thoughts about services, goods, and events has grown due to user-generated content on websites and social networks like Twitter. Decision-makers are utilizing these networks more [10]. Sentiment analysis systems may reveal people's thoughts and views. By evaluating these thoughts, we may determine the users' motivations for social problems' success or failure.

We can conclude from the literature that these models are excellent in training data. due to their increased features in the test data, they confront issues such as overfitting and low performance in Sentiment analysis and text processing. Sentiment analysis suffers from an uneven data set issue. Feature selection methods are always advised to address the high dimensionality-related issues. A calculation scheme is used in the feature selection process to remove unnecessary and redundant features from the high-dimensional feature space. Thus, only the most informative features are selected. Therefore, the things mentioned in this article, whose structure is depicted in Figures 1 and 2, have been successfully implemented. The following is a list of the contributions made by the suggested method:

i. To analyze opinions from text comments, develop an automated text analysis system.
ii. Analyzing the effects of enhancing sentiment categorization by employing the BCA and OCA feature selection approaches.
iii. To increase classification accuracy, avoid overfitting and overlap, and reduce storage complexity, another set of studies was carried out using BCA and OCA.
iv. Using the original feature set and the result of feature selection to apply three well-known supervised deep learning models, including RNN, CNN, and hybrid (CNN-RNN).

II. RELATED WORKS

According to the quality of the sentiment analysis, Zainuddin et al. provide a Principal Component Analysis (PCA) based feature selection method that can zero in on the best features to use in the process. Feature selection helps get rid of extra details that can lower classification precision. This research introduces a feature selection strategy for labeling Twitter emotions PCA-based on PCA. Support Vector Machine (SVM) classification is improved by combining PCA with the dictionary-based Sentiwordnet approach. This contrasts the 94.53% and 97.93% achieved in experiments using the STS benchmark and the HCTS dataset. The proposed strategy in this study exhibits promising results in enhancing sentiment analysis performance compared to previous statistical feature selection methods [11].

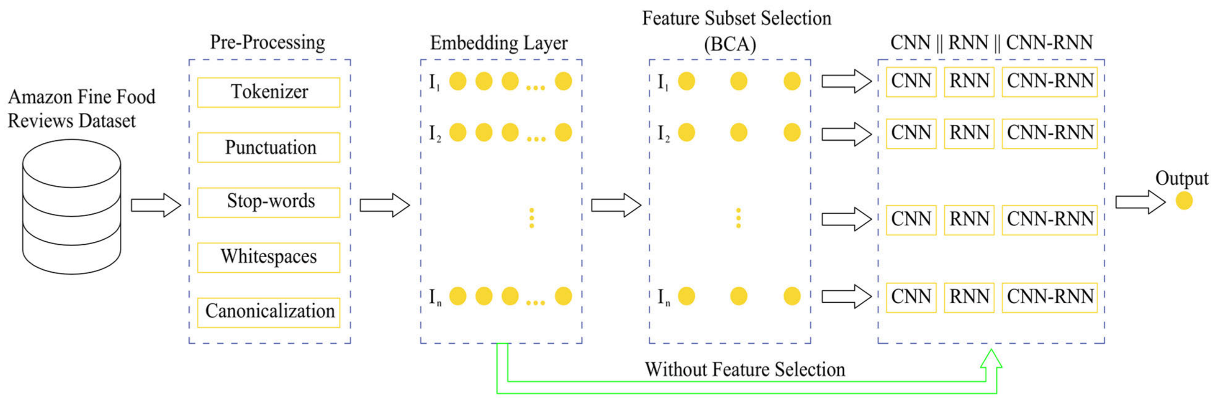


FIGURE 1. Proposed model.

```
Inputs: D: Raw Amazon Fine Food Reviews dataset, comment: a comment which is selected from the D;
For comment=1:D   // Preprocessing steps
        comment=Delete stop-words, sparse terms, specific words of comment
        comment= Accent marks before lemmatizing the text data of comment
        comment=Remove accent marks, punctuation, and diacritics of comment
End for
CNNM= Build a CNN model which sentiment comment based on CNN architecture
RNNM= Build a RNN model which sentiment comment based on RNN architecture
CNN-RNNM= Build a hybrid model include CNN and RNN models which sentiment comment based on CNN and RNN architectures
[accuracy_CNNM, precision_CNNM, recall_CNNM, f1-score_CNNM] = Calculate average metrics [Accuracy, Precision, Recall, F1-score] based on five-fold CVs from CNNM model
[accuracy_RNNM, precision_RNNM, recall_RNNM, f1-score_RNNM] = Calculate average metrics [Accuracy, Precision, Recall, F1-score] based on five-fold CVs from RNNM model
[accuracy_CNN-RNNM, precision_CNN-RNNM, recall_CNN-RNNM, f1-score_CNN-RNNM] = Calculate average metrics [Accuracy, Precision, Recall, F1-score] based on five-fold CVs from CNN-RNNM model
[AAT_CNN, AAT_RNN, AAT_CNN-RNN] = Calculate Average-Time-Test from [CNNM, RNNM, CNN-RNNM] models
For comment=1:D   // Feature Selection step
        commentF= Select features from comment based on BCA technique
End for
CNNFM= Build a CNN model which sentiment commentF based on CNN architecture
RNNFM= Build a RNN model which sentiment commentF based on RNN architecture
CNN-RNNFM= Build a hybrid model including CNN and RNN models which sentiment commentF based on CNN and RNN architectures
[accuracy_CNNFM, precision_CNNFM, recall_CNNFM, f1-score_CNNFM] = Calculate average metrics [Accuracy, Precision, Recall, F1-score] based on five-fold CVs from CNNFM model
[accuracy_RNNFM, precision_RNNFM, recall_RNNFM, f1-score_RNNFM] = Calculate average metrics [Accuracy, Precision, Recall, F1-score] based on five-fold CVs from RNNFM model
[accuracy_CNN-RNNFM, precision_CNN-RNNFM, recall_CNN-RNNFM, f1-score_CNN-RNNFM] = Calculate average metrics [Accuracy, Precision, Recall, F1-score] based on five-fold CVs from CNN-RNNFM model
[AAT_CNNF, AAT_RNNF, AAT_CNN-RNNF] = Calculate Average-Time-Test from [CNNFM, RNNFM, CNN-RNNFM] models
End
```

FIGURE 2. Pseudocode of the proposed algorithm.

Long short-term memory (LSTM) [12] validation takes longer, although the time is reduced if you have previously selected the features. For example, to train an LSTM model for a million data records, a pre-feature selection process can help us to obtain a more organized data. As a result, the data training will be more accessible on the model. Moreover, Changes to the settings and weights don't take long, and excellent outcomes can be achieved.

The demand for collecting and analyzing social media data has expanded over the past few decades. Twitter is an excellent place to find this information. Sentiment Analysis can employ user reviews of various events or subjects in words of sentiment texts for decision-making. However, tweets are unstructured data and can contain millions of records when collected. As part of the data analytics process, the raw data must be transformed into a new representation known as features. Due to duplicate and irrelevant features, the number of features in a large dataset might easily exceed 100,000 in dimensionality. The performance of machine learning algorithms is adversely affected and degraded when applied to the high-dimensional feature space. Additionally, the high dimensionality of feature space results in overfitting the learning algorithm, which increases the learning time and memory required for data analytics [6]. Methods based on machine learning and deep learning algorithms have been used in these endeavors. It's been found that deep learning outperforms these two methods even if they have yielded promising results [13]. Based on emotion, monkeypox tweets have been classified and displayed [14]. They collected more than 800,000 tweets from Twitter in 103 different languages. The dataset underwent preprocessing, NRCLex labeling, and the identification of eight emotions for classification. In order to obtain accuracy and prevent bias in the model, Synthetic Minority Oversampling Technique (SMOTE) Oversampling and Random Undersampling methods were utilized to address the data imbalance in the dataset. Additionally, they created the One-Dimensional Convolutional Neural Networks (1DCNN), LSTM, Convolutional LSTM (C-LSTM), and Bi-directional LSTM (BiLSTM) models, which are four acceptable models for emotion classification based on deep learning architecture. Based on our analysis, they discovered that the models used in this research can accurately predict emotions by 95%–96%. Using 84,018 tweets on the monkeypox virus, a hybrid CNN-LSTM model was utilized in a related work [15] to perform sentiment analysis and identify the emotional polarity of the material in the tweets. Based on user opinion of the monkeypox virus as positive, negative, or neutral, the findings of this study were obtained. The study's model performed well on average, averaging 83% accuracy, thanks to the optimizers. This method is deemed to be somewhat effective and appropriate for categorizing feelings in tweets about the monkeypox virus based on the other metrics obtained.

[16] proposed using collaborative filtering on online social networks in conjunction with hybrid deep-learning models to apply sentiment analysis to recommender systems. The system architecture can combine several strategies to implement suggestions, such as sentiment analysis methods, hybrid deep-learning models, and preprocessing techniques. Two public datasets, including Amazon Fine Foods Reviews and Amazon Movie Reviews, are utilized to test the proposed method. The main objective of [17] was to develop a system that can classify each input review as positive or negative sentimentality by comparing the effectiveness of three widely used supervised learning classifiers: Naive Bayes (NB), Logistic Regression (LR), and SVM. They used Twitter data about OK cuisine reviews from the Amazon website to analyze sentiment. It turns raw data into helpful information for many research fields and identifies patterns to predict future trends in the research domain.

A novel method, based on chi-square, has been implemented by Paudel et al. for choosing the right number of features to include, and those features are picked out using the standard score threshold. It has been discovered that the logarithm of the number of features chosen correlates with the result of the sentiment classification when using learning-based techniques. This connection exists independently of any learning procedure. This study adds to the growing body of evidence indicating that researchers may, in many cases, select the optimal number of features in learning-based approaches to produce the highest performance in sentiment classification. Researchers will benefit from this as they better hone in on the traits that will maximize the effectiveness of algorithms that rely on learning [18].

To solve the scalability problems caused by ever-growing feature sets in sentiment analysis, Iqbal et al., presented a hybrid solution based on a Genetic Algorithm (GA) based strategy that can reduce feature set size by up to 42% without sacrificing accuracy. They looked examined two additional popular methods of feature reduction and how they stacked up against their suggested GA-based method: PCA and Latent Symantec Analysis (LSA). Both PCA and LSA were beaten by the GA-based technique, with the former showing a 15.4 percentage point improvement and the latter showing a 40.2 percentage point improvement. SentiWordNet, Machine Learning, and Machine Learning with GA optimal feature selection have all been utilized for sentiment analysis. The SWN method's accuracy is consistently worse than the other two approaches we looked at (56% best case). Their GA-integrated model achieves a 36%-43% reduction in feature size and a 5% improvement in performance compared to the ML method. They tested the effectiveness of their suggested model using six distinct classifiers, including (J48, NB, PART, SMO, IB-k, and JRIP). Using GA-based feature selection, the NB classifier achieved the maximum accuracy (about 80%) on the Twitter and feedback dataset. On the contrary, IB-95% k's accuracy in the geopolitical dataset was the best of any classifier tested [19].

SVM, NB, decision tree, and K-nearest neighbor are the four primary techniques used for sentiment analysis that Wazery et al. implemented. A deep neural network (RNN) was used that makes use of a long short-term memory mechanism. Two methods were applied to three different

Twitter datasets (including those from IMDB, Amazon, and an airline). A comparison of these algorithms, with experimental findings showing that the RNN employing LSTM achieves the highest accuracy (88, 87, and 93%, respectively) [20]. Through real-time data extraction, Sharma et al. evaluated tweets. Text data underwent preprocessing feature extraction, the Chi-square test, PCA, and several machine learning classifications (SVM, NB, Random Forest, and LR) against key performance indicators [21].

Another use for sentiment analysis is in the medical profession. [22] used a questionnaire survey to determine if people who received the Corona immunization would recommend it to others. The Bidirectional Encoder Representations from Transformers (BERT) approach and the Valence Aware Dictionary for sEntiment Reasoning (VADER) technique were used to examine all of the comments. The accuracy of the findings is between 71 and 75 percent. The remainder of the paper is structured as follows: Section III explains the methodology. The main results and discussion are presented in Section IV. In section V, the main ideas of this essay are concluded. Table 1 displays several published studies on the subject of sentiment analysis.

TABLE 1. A summary of previous studies.

Year	Ref	Methods	Data Numbers	Acc (%)
2016	[11]	PCA and SVM	STS and HCTS	94.53 and 97.93
2023	[14]	1-DCNN, LSTM, C-LSTM	800,000 tweets	95-96
2023	[15]	hybrid CNN-LSTM	84,018 tweets	83
2021	[16]	L-CNN and C-LSTM	Amazon Fine Foods Reviews	80.04 and 79.95
		L-CNN and C-LSTM	Amazon Movie Reviews	82.27 and 82.27
2019	[19]	GA with NB	Twitter and feedback	80
2018	[20]	RNN with LSTM	IMDB	93
2022	[22]	BERT and VADER	725 comments	71 and 75

III. METHODOLOGY

The proposed model consists of six stages, which are data collection, preprocessing, embedding Layer, BCA, deep learning techniques, and sentiment analysis. Figure 1 and Figure 2 show the structure and the pseudocode of the proposed method respectively. It was necessary to perform some preliminary processing on the raw data to make it more machine-readable. As a result, the network can only understand numbers now that word embedding has been employed to convert words to their associated vectors. Data will be input into the BCA algorithm for feature selection after the preceding operations have been completed. Deep learning techniques will be used to train the data after that. Finally, the model can be saved and utilized for sentiment analysis and classification. The remaining four sections of the study are organized as follows: Dataset, Text Pre-processing, Word Embedding, Feature Selection Algorithm, Feature subset Selection Based on Binary Coordinate Ascent, Deep Learning techniques, Classification Techniques, and Evaluation Criteria.

A. DATA COLLECTION

Our study specifically employed data from Amazon Fine Food Reviews, a collection of 568,454 reviews; in this study, we used all the opinions in the dataset to examine users' opinions. The reviewers' product IDs and comments on the products are contained in a single CSV file. Besides, there are also the reviewer's ratings (ranging from 1 to 5). In addition to the content of the reviews, a timestamp is attached to each review [16]. For our labels and raw inputs, we extract the scores and review texts from the scores and review texts. The collected dataset is imbalanced. It consists of 82,037 negative comments and 486,419 positive comments, indicating a significant class imbalance. This information is important to consider when assessing the performance and generalizability of the deep learning model. In the dataset, 80% of the data is used for training, and the remaining 20% is used for testing.

B. PREPROCESSING

It is essential to utilize text pre-processing procedures to convert the natural language to a machine-readable format. In this step, all letters and numbers in the text data are converted to lowercase or uppercase letters and words. Text data can be cleaned of numbers that aren't relevant to text analysis or don't lead to valuable data products. Accent marks are also stripped of punctuation, as is the space between textual data, stop-words, dispersed terms, and specified terms. Before any text can be used, it must first be processed in some way. Text data must be normalized after it has been prepared for analysis. Text pre-processing necessitates the use of the following methods [10]:

1. Convert all letters in text data to lowercase letters or uppercase Letters.
2. Convert numbers to words or remove numbers from text data.
3. Delete punctuation, accent marks, and diacritics.
4. Clear Whitespaces of text data.
5. Expanding abbreviations.
6. Delete stop-words, sparse terms, and particular words.
7. Canonicalization of text data.

C. EMBEDDING LAYER

In attribute extraction, word vector embedding is the first step. The Word2Vec model will be used in this case. The Word2Vec model is used to convert tweets into the appropriate n-dimensional vectors. The embedding vector is used by copying the weights to the embedding layer. This layer serves as a conduit for the input to the network (with any structure). Word embedding is another name for this method. Word embedding is based on the idea that all of a language's words can be represented numerically (in vector form). Embedding is an n-dimensional word processor that interprets the meaning and content of words using numerical numbers. No of how useful a given "word vector" may or may not be, each collection of numbers is considered to have been included. In this collection, we've gathered word vectors that our apps can use to understand better words and the relationships between them, as well as the content of individual words [23].

It uses a neural network to display all words and texts on a fixed vector. This vector is calculated using the correct numbers during the model's training. Instead of displaying a specific word or characteristic, each column in this vector displays a number. Each word will have a unique representation in a 400-dimensional space if we estimate this vector to be 400. It is also necessary that the primary data set for training the model include billions of words contained in millions of documents or texts in order to improve its accuracy. A text or news item's vector can be found by creating vectors for each word in the text or article. The average of the numbers in each column can also be determined for each text or document, resulting in a vector.

D. FEATURE SUBSET SELECTION

There are many features, but most are either not used or don't require significant data. If you leave these features in place, there won't be any informational issues, but the network will have to work more to analyze sentiment. Additionally, it enables us to store both valuable and a lot of meaningless info. The feature selection problem has received numerous proposals for solutions and methods.

Figure 3 (a) shows a general feature selection framework that illustrates how the algorithm can lower your model's input variable by using only relevant data and eliminating noise. In this method, a machine learning model has its characteristics chosen automatically based on the type of work being done. Figure 3 (b) based on the values of the objective function, the BCA algorithm iteratively adds and eliminates features from the chosen subset of characteristics. The BCA verifies, at each cycle, if the presence of a feature in a particular collection of characteristics increases or decreases the classification performance. Let's pretend that a feature was inadvertently added or subtracted from the feature subset. In that circumstance, the BCA algorithm is at liberty to fix the incorrect decisions through subsequent BCA scans, coming as close as possible to the ideal result.

On the other hand, the abundance of features lowers the test data's classification results' accuracy. In other words, overfitting is a concern. Overfitting occurs when the learning model does not generalize to test data and has a significant error in the test data. When there is a high correlation between the independent variables, it raises questions about the model's validity despite the high value of the correlation

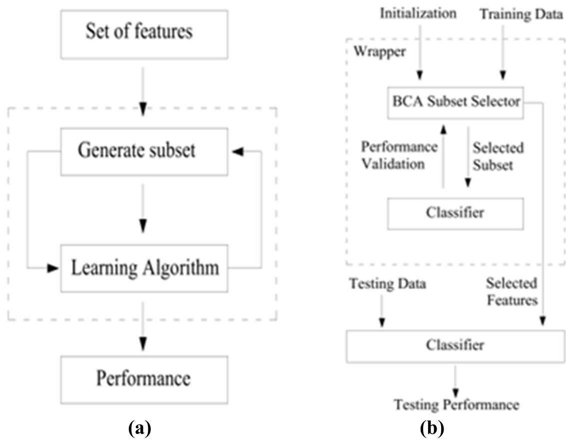


FIGURE 3. Feature selection structure, a) A general framework of feature selection, b) BCA feature subset selection.

coefficient between the dependent and independent variables. That is to say, the model is superficially appealing, but it is missing critical, independent variables [24]. The BCA and OCA techniques have been applied to choose useful features and address this issue. Evolutionary algorithms can be used to address the NP-Hard problem of choosing beneficial characteristics for categorization.

1) BINARY COORDINATE ASCENT (BCA)

The Binary Coordinate Ascent (BCA) algorithm is a deterministic method for iteratively finding the best solution in a given region. In 2016, Zarshenas et al. published it. Older feature selection methods like Sequential Feature Selection and Sequential Forward Floating Selection, two of the most well-known wrapper-based feature subset selection (FSS) algorithms, need a tremendous amount of subset evaluations, which is why BCA is so important (IWSSr). The BCA method starts its search at a point in space where the binary representation of input variables yields continuous output values to identify the best set of features based on their utility. The provided cost function is only iterated once at each point when BCA changes the solution. The BCA algorithm helps machine-learning algorithms choose the best or worst set of features from the feature space to carry out a particular task, like classification, as well as efficiently. The BCA evaluates whether a specific feature inside a specific group of characteristics improves or lowers classification performance at each iteration. What transpires if a feature is inadvertently included or excluded from the feature subset? To get as near the optimal solution as possible, BCA will seek to repair inaccuracies caused by earlier scans. As a result, BCA dramatically reduces the time needed to run the model and improves its precision. Figure 3 displays the overall block diagram for BCA [25].

2) OPTIMAL COORDINATE ASCENT (OCA)

Optimal Coordinate Ascent (OCA), a novel technique that enables choosing features from both blocks and individual features. To get the best solution for the gradient

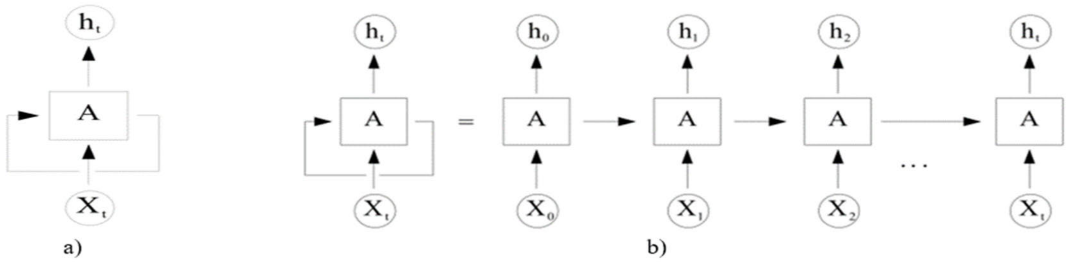


FIGURE 4. **RNN structure a) Simple RNN, b) Opened loop of RNN.**

boosting technique score (the number of properly categorized samples), OCA uses coordinate ascent. The idea of interdependencies between variables constituting blocks in optimization is taken into consideration by OCA. The Natural Processing (NP) hard original problem, where the number of options quickly increases and renders a grid search impractical, is resolved using coordinate ascent optimization [26], [27]. The number of rounds required to convert this NP-hard issue into a polynomial search problem has been significantly reduced.

E. DEEP LEARNING TECHNIQUES

In deep learning, a particular branch of machine learning, software trains by subjecting multilayer neural networks to enormous volumes of data. Excellent NLP outcomes, including sentiment analysis across different datasets, have been produced by deep learning advancements [28]. The most significant benefit of deep learning is the elimination of the requirement for manual feature extraction. Instead, they use word embedding as input that contains context information, and during the training phase, the middle layers of the neural network self-learn the features. Words are encoded in a high-dimensional vector, and a neural network extracts their features. Deep learning gets going quickly due to its higher performance across a range of problems and the fact that it is autonomous, which makes problem-solving much more pleasant. A deep network of simple data transformations (layers) is used to assign inputs to targets, and these layers learn by observing many input and target examples. Transformation is carried out through a layer that is parameterized by one's weights. For the network to properly arrange the input samples for the associated targets, learning layers entails finding a series of values for the weights of each layer. The deep neural network may include millions of parameters, and selecting the appropriate value for each parameter appears arduous. The behavior of all other parameters will change if the value of one parameter is changed [29].

In sentiment analysis, classification refers to grouping information into different categories by the equation used to determine the sentiment of the text. For binary classification (positive and negative) of the Amazon Fine Food Reviews dataset, our study used three techniques: RNN, CNN, and hybrid (CNN-RNN). The following is a description of the study's final three sections: RNN, CNN, and hybrid neural network techniques.

1) RECURRENT NEURAL NETWORKS (RNNs)

An artificial deep neural network is known as an RNN. Many NLP research employs this technique. They are made to identify a sequence of data features. RNNs were first developed in 1980; however, they have recently become trendy. The leading causes are the general advancements in neural network architecture and the vast increase in computational power, particularly the effectiveness of graphics cards' parallel processing units. As demonstrated in Figure 4(a), these neural networks are effective for processing confidential or sequential data because each processing unit (neurons) may keep an internal state (memory) that preserves the information from prior input. This capability is essential for several applications using confidential data [30].

This function aids the network in maintaining the internal state necessary for comprehension and connection-finding between various words in longer sequences. It should be mentioned that we interpret a sentence's meaning based on the context in which each word is used when we read it. Utilizing this data series structure is the fundamental goal of this form of architecture. This neural network's name is derived from the fact that recursive networks of this kind operate. In other words, this operation is carried out for each word or sentence in a sequence, and the results rely on the input and prior operations. This is accomplished by repeating one of the network's outputs at time t-1 with its input at time t (i.e., combining the output from the previous step with the fresh input from the current step) [31]. These cycles enable data entry between one time step and the following time step. In other words, these networks contain a loop that allows them to transmit information to neurons while simultaneously reading input. Figure 4(b) shows the opened loop of RNN.

For the RNN to make an accurate prediction, it needs data from the beginning of the phrase. These dependencies are regarded as long-term dependencies since the time between the pertinent data and the moment it is used to construct a forecast is usually rather considerable. Unfortunately, learning these dependencies becomes increasingly challenging in practice as distance increases because they either have the problem of gradient vanishing or explosion.

These problems happen during deep network training when gradients are propagated from the end to the beginning of the network. Since the gradients from the end levels had to go through numerous multiplications to reach the starting layers, their values began to decrease (less than 1). As the process continues, these values decrease to the point where (in extremely deep networks) they are so minor that they terminate the training process since the gradients are so small that they have no impact on the weight changes. The term "gradient disappearing" refers to this issue. Similar to this, there is a chance for gradient explosion, in which the gradient's values progressively grow to such a size that the model finally makes a mistake, resulting in an overflow of the computations. By exposing the network, we can more clearly grasp the nature of the loop in this type of system. Figure 5 illustrates a detail of RNN's cell [32].

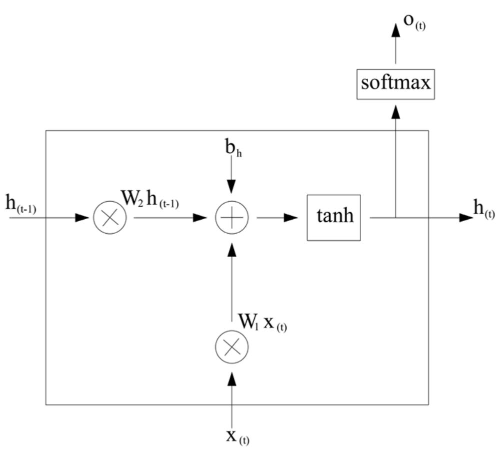


FIGURE 5. Details of an RNN's cell.

2) CONVOLUTIONAL NEURAL NETWORKS (CNNs)

CNN is a particular type of neural network initially designed for computer vision and uses layers with filtering that implements local characteristics. It has many applications, including computer vision, speech recognition, and NLP. The network includes neurons with weights and biases altered based on training data by a learning algorithm. It also has local receptive fields, which are discrete areas of input layer neurons coupled to hidden layer neurons. Convolutional layers, a pooling layer, and one or more fully linked layers make up the CNN structure [24]. Figure 6 illustrates how 1d-CNN, which Kim designed initially, works with patterns in one dimension and has the propensity to be helpful in natural language processing. It accepts phrases of various lengths as input and outputs fixed-length vectors [25]. The longest sentence that the network can process is clipped, and the shortest sentence is filled with zero vectors. The dropout regularization used to control over-fitting follows [33].

3) HYBRID NEURAL NETWORK TECHNIQUES

Integrating CNN and RNN neural networks, CNN-RNN is a hybrid neural network technique. Each model is distinct in its structure. To achieve its aim of extracting the most local properties, CNN's convolutional layer employs a variety of filters with different window sizes. A pooling layer is then implemented to enhance feature extraction by concentrating on the most crucial ones. The retrieved attributes from the following layer also have a lower number. The data are converted into one-dimensional feature vectors using a flattened layer before being sent as inputs to the RNN. The RNN then gets the features that CNN has retrieved. Combining these two models provides a model with the benefits of both CNN and RNN [33].

F. EVALUATION CRITERIA

The metrics in this section are used to assess the output of CNN, RNN, and hybrid (CNN-RNN) models trained on the Amazon Fine Food Reviews dataset with and without the BCA method. Metrics for precision, accuracy, recall, and f1-score are used to compare the outcomes. Evaluation of the performance of any deep learning model is a crucial step in its development. Each technique's performance is calculated by computing various metrics. The main reason for utilizing alternative metrics is to determine how well a deep learning model can function with unlabeled data. The following measures are used in this paper [28]:

1. Accuracy is the ratio of correctly analyzed samples to the overall sample count. This is to evaluate the accuracy of the model. The ratio of accurately predicted reviews to the total review under examination is used to calculate accuracy. It can be calculated using equation (1).

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{T_P + T_N}}{\boldsymbol{T_P + T_N + F_P + F_N}} \tag{1}$$

2. Precision based on equation (2) aimed at evaluating model precision. It is determined by dividing the number of correctly predicted reviews by the number of correctly predicted reviews (TP+FP), with the caveat that a contained class (c) may contain either positive or negative reviews.

$$\boldsymbol{Precision} = \frac{\boldsymbol{T_P}}{\boldsymbol{T_P + F_P}} \tag{2}$$

3. Recall: It is measured as the ratio of TP to the number of actual reviews (TP+FN) and having a negative or positive class. Through that, the model completeness is measured. It can be calculated using equation (3).

$$\boldsymbol{Recall} = \boldsymbol{T_P / (T_P + F_N)} \tag{3}$$

4. F1-score: calculates the test's recall and precision to determine the score. The following is the formal definition of f1-score:

$$\boldsymbol{F_1 score} = 2 * \frac{\boldsymbol{Precision * Recall}}{\boldsymbol{Precision + Recall}} \tag{4}$$

5. TP is the true positive and was accurately anticipated in the equations above. A TN is one that was accurately predicted, while a false positive is one that was wrongly forecasted. FN stands for false-negative and erroneously predicted.

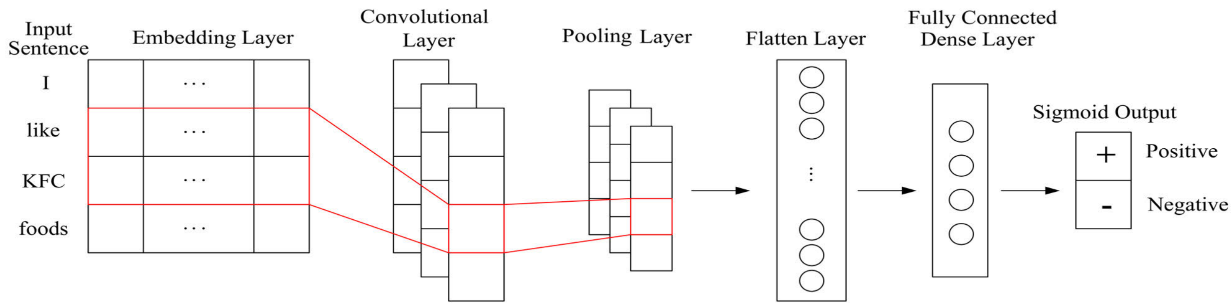


FIGURE 6. The structure of CNN.

IV. RESUTLTS AND EVALUATION

Figure 1 illustrates the several phases of the proposed system model for classifying Twitter sentiment. These phases, as mentioned in Figure 2, including reading the data, preprocessing the data, and extracting the features. Next, a feature subset is chosen based on BCA, and the classification phase is performed. The proposed system accomplishes classification in two ways: classification of the original feature space and classification of features chosen by the BCA technique using a wrapper. Finally, the accuracy, recall, F1 score, and precision of the three classification algorithms (CNN, RNN, and hybrid CNN-RNN) have been compared.

Python programming language based on Keras's sequential model was employed for implementing the RNN, CNN, and hybrid (CNN-RNN) models in this work. All training processes were conducted via an NVIDIA GeForce GTX 1080 Ti GPU.

Training and testing are all included in this section, after which the outcomes are assessed using RNN, CNN, and hybrid CNN-RNN models. The following parts go into further information on it.

A. TRAIN AND TEST PHASES

Here are some instances of settings in which each approach produces the desired effect. To prevent the neural network from becoming unduly dependent on the training data, each model contains a dropout layer with a rate of 0.5. The models were built using the Adam optimizer [35], with a learning rate of 0.001, a batch size of 128 for 15 epochs, and a fully-connected dense layer with sigmoid activation [36] that produces a binary prediction. The CNN model contains three layers of 1d-CNN that are all implemented with 64 filters and kernel sizes of 1, 2, and 3, respectively. After each layer, a max-pooling layer [37] with two pooling filter sizes is applied, which chooses just the value with the greatest weight and ignores the remaining values, considerably improving the convolutional layer's results and reducing the input to the next layer. It also contains a flattening layer that converts a two-dimensional feature matrix into a vector that can be fed into the output layer. The CNN-RNN model contains one layer of 1d-CNN with 64 filters and a kernel size of 1, followed by a max-pooling layer with two pooling filter sizes, and one layer of RNN with 64 neurons in each layer.

The cross-validation parameter of the BCA feature selection is set to 5, which is set as cross_val_score, and the delta with 1e-05 rate is used to assess the convergence of the objective function. Scoring equal to accuracy is used as the metric to be used as the objective to be maximized.

RNN, CNN, and hybrid CNN-RNN deep learning architectures were used to train the proposed sentiment analysis approach. To illustrate the system's performance, we used fivefold cross-validation (CV) [38] to train and test our models five times with different data, utilizing 80% of the data for training and 20% for testing. The average of all five-fold CVs was computed. Results from all types are shown in Table 2.

B. RESULTS

Three models of RNN, CNN, and a hybrid CNN-RNN deep learning architecture are shown here, along with experimental results. Results from the proposed sentiment analysis are presented and discussed first. Next, the model's ATT (average test time) during the test is calculated.

The average performance of three different deep learning architectures using the fivefold CV applied to all sentiment assessments is shown in Table 1. According to the data, all of our models achieve an accuracy of at least 73%, with the Hybrid CNN-RNN with BCA achieving the highest accuracy.

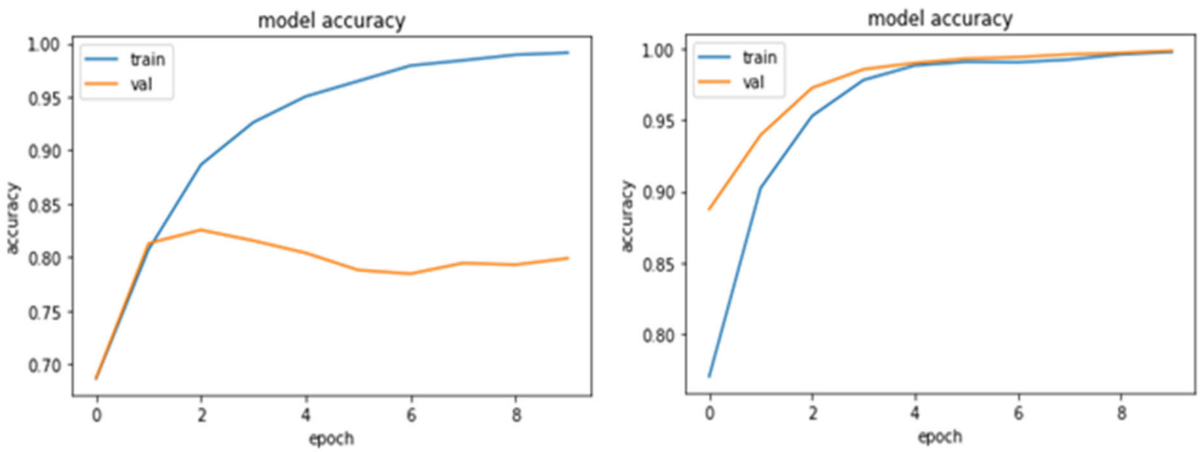
Prediction accuracy for the hybrid (CNN-RNN) model trained on the Amazon Fine Food Reviews dataset was 74.28% before combining with feature selection methods, and it has grown to 97.91% after combining with the BCA approach, as shown in Table 2. Combining CNN and RNN models in the BCA method improves prediction accuracy to 96.24% and 96.21%, respectively. When the OCA feature selection approach is utilized, our performance model's accuracy increases by 12 to 13 percentage points and reaches up to 85%. Despite the fact that BCA outperforms OCA, applying feature selection methods to our suggested model boosted model accuracy.

Based on an analysis of the combined model results, we find that the BCA method produces the highest accuracy (97.91%) when applied to the Amazon Fine Food Reviews dataset. Figure 7 illustrates how, during the initial stages of model training, the accuracy of system performance can be observed and how feature selection techniques can be used to increase this accuracy. In this study, the model's accuracy

TABLE 2. The performance of the proposed system in comparison with state-of-the-art techniques.

Techniques	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ATT (s)*
L-CNN [14]	80.04	-	-	80.24	-
C-LSTM [14]	79.95	-	-	80.00	-
SVM with TF-IDF [15]	91	93	91	91	5.8271
NB with TF-IDF [15]	81	91	81	87	1. 8554
LR with TF-IDF [15]	88	91	88	90	1.4192
CNN	73.36	73.91	73.22	73.46	0.0975
RNN	73.15	73.90	73.51	73.49	0.103
Hybrid (CNN-RNN)	74.28	74.81	74.22	74.41	0.183
CNN with OCA	84.03	84.01	84	84.08	0.172
CNN with BCA	96.24	96.45	96.31	96.29	0.154
RNN with OCA	84.11	84.31	84.30	84.27	0.201
RNN with BCA	96.61	96.60	96.69	96.60	0.197
Hybrid (CNN-RNN) with OCA	85.41	84.58	84.50	84.71	0.271
Hybrid (CNN-RNN) with BCA	97.91	97.90	97.64	96.19	0.219

*ATT = Average Time Test



(CNN-RNN) without BCA algorithm (CNN-RNN) with BCA algorithm

FIGURE 7. The accuracy of the models before and after the BCA used the Amazon Fine Food Reviews dataset.

increased to over 90% within the first 5 epochs of training, demonstrating the benefit of applying feature selection approaches to improve the accuracy of the models from the outset of the model training process.

Table 2 also shows that the employment of feature selection techniques significantly increases accuracy, indicating that feature selection techniques can improve the accuracy of sentiment analysis performance by causing the selection of relevant features. The use of the BCA technique improved system performance by a considerable 23 percent, from 73% to 96%, based on CNN and RNN models, in this paper. When employing CNN and RNN models to analyze people's opinions, reducing the complexity of data is an advantage of applying feature selection techniques. Using comments reduces the complexity of data and improves the system's performance accuracy. However, despite spending more time training and testing, the hybrid models utilized in this article have significantly improved the system's performance, illustrating the importance of hybrid models in sentiment analysis.

Comparing the results obtained in the paper [16], which is based on two hybrid models (L-CNN and C-LSTM) to our proposed results (which are based on hybrid models with or without the use of BCA) on the Amazon dataset, it is clear that the accuracy of our model without using the feature was insufficient; however, when the BCA was used in the proposed model, the performance of the system quickly increased, once again demonstrating the significance of feature selection to overall system efficiency.

For sentiment analysis, the researchers in this work [17] employed machine learning algorithms such as SVM with

TF-IDF, NB with TF-IDF, and LR with TF-IDF; even with a lengthy testing period, the system achieved an accuracy of 91%. In comparison, our proposed model's performance accuracy reached 97% with a significantly shorter ATT. The results demonstrate the powerful effectiveness of deep learning methods combined with feature selection algorithms.

Another objective of this research has been to take into account the ATT. It's critical to assess a deep learning model's efficiency in addition to its accuracy when evaluating its performance. The model's processing time for a single input may be measured using the average time test. This data is useful for assessing the real-time capacity of the model and its applicability for certain applications. Table 2 displays the results of our calculation of all the ATT resulting from testing models. The whole computed ATT is less than 1 second, and it is drastically lower when compared to [17]. The ATT time for the ideal model, which is the Hybrid (CNN-RNN with BCA), which is a shorter time, is 0.219 second. The smallest amount of time when the model is evaluated by CNN is 0.0975 second.

V. CONCLUSION

In this paper, we present a combination of feature subset selection algorithms based on the BCA with the CNN, RNN, and hybrid (CNN-RNN) models, which increases all models' computational accuracy. In explaining this result, it can be said that all models usually have good training data accuracy, but the models' accuracy decreases in testing data. This problem has been solved by selecting the BCA subset feature by choosing the features that are highly dependent on the response and increasing the accuracy of all models. This method increases the probability of correctly predicting the class's occurrence by selecting the independent category (predictor) feature that correlates highly with the dependent category (response) feature. However, the method presented in this framework has unavoidable problems such as generalizing the results based on other datasets and searching across the feature subsets required in the subset construction step, which needs its computational complexity. Also, results show that:

1. The proposed model obtained higher accuracy than three state-of-the-art recurrent approaches;
2. BCA is more effective in increasing accuracy in sentiment analysis;
3. Implementing the network, an imitation strategy shows that our proposed model is strong for text classification.

REFERENCES

[1] S. Khanna, B. Tiwari, P. Das, and A. K. Das, "A comparative study on various text classification methods," in *Proc. Comput. Intell. Pattern Recognit. (CIPR)*. Singapore: Springer, 2020, pp. 539–549, doi: 10.1007/978-981-15-2449-3_46.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[3] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6645–6649, doi: 10.1109/ICASSP.2013.6638947.

[4] S. Liao, J. Wang, R. Yu, K. Sato, and Z. Cheng, "CNN for situations understanding based on sentiment analysis of Twitter data," *Proc. Comput. Sci.*, vol. 111, pp. 376–381, Jan. 2017, doi: 10.1016/j.procs.2017.06.037.

[5] J. L. Elman, "Finding structure in time," *Cogn. Sci.*, vol. 14, no. 2, pp. 179–211, Mar. 1990, doi: 10.1207/s15516709cog1402_1.

[6] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *WIREs Data Mining Knowl. Discovery*, vol. 8, no. 4, Jul. 2018, Art. no. e1253, doi: 10.1002/widm.1253.

[7] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, Dec. 2014, doi: 10.1016/j.asej.2014.04.011.

[8] S. Akyol and B. Alatas, "Sentiment classification within online social media using whale optimization algorithm and social impact theory based optimization," *Phys. A, Stat. Mech. Appl.*, vol. 540, Feb. 2020, Art. no. 123094, doi: 10.1016/j.physa.2019.123094.

[9] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: A review," *Artif. Intell. Rev.*, vol. 53, no. 6, pp. 4335–4385, Aug. 2020, doi: 10.1007/s10462-019-09794-5.

[10] M. Hussein and F. Özyurt, "A new technique for sentiment analysis system based on deep learning using chi-square feature selection methods," *Balkan J. Electr. Comput. Eng.*, vol. 9, no. 4, pp. 320–326, Oct. 2021, doi: 10.17694/bajece.887339.

[11] N. Zainuddin, A. Selamat, and R. Ibrahim, "Twitter feature selection and classification using support vector machine for aspect-based sentiment analysis," in *Proc. 29th Int. Conf. Ind. Eng., Other Appl. Appl. Intell. Syst. Trends Appl. Knowl.-Based Syst. Data Sci. (IEA/AIE)*. Morioka, Japan: Springer, Aug. 2016, pp. 269–279, doi: 10.1007/978-3-319-42007-3_23.

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[13] M. H. Abdalla and M. Karabatak, "To review and compare evolutionary algorithms in optimization of distributed database query," in *Proc. 8th Int. Symp. Digit. Forensics Secur. (ISDFS)*, Jun. 2020, pp. 1–5, doi: 10.1109/ISDFS49300.2020.9116418.

[14] R. Olusegun, T. Oladunni, H. Audu, Y. Houkpati, and S. Bengesi, "Text mining and emotion classification on monkeypox Twitter dataset: A deep learning-natural language processing (NLP) approach," *IEEE Access*, vol. 11, pp. 49882–49894, 2023, doi: 10.1109/ACCESS.2023.3277868.

[15] O. Iparraguirre-Villanueva, A. Alvarez-Risco, J. L. H. Salazar, S. Beltozar-Clemente, J. Zapata-Paulini, J. A. Yáñez, and M. Cabanillas-Carbonell, "The public health contribution of sentiment analysis of monkeypox tweets to detect polarities using the CNN-LSTM model," *Vaccines*, vol. 11, no. 2, p. 312, Jan. 2023, doi: 10.3390/vaccines11020312.

[16] C. N. Dang, M. N. Moreno-García, and F. D. L. Prieta, "An approach to integrating sentiment analysis into recommender systems," *Sensors*, vol. 21, no. 16, p. 5666, Aug. 2021, doi: 10.3390/s21165666.

[17] F. H. Fadel and S. F. Behadili, "A comparative study for supervised learning algorithms to analyze sentiment tweets," *Iraqi J. Sci.*, pp. 2712–2724, Jun. 2022, doi: 10.24996/ijs.2022.63.6.36.

[18] S. Paudel, P. W. Prasad, A. Alsadoon, M. D. R. Islam, and A. Elchouemi, "Feature selection approach for Twitter sentiment analysis and text classification based on chi-square and Naïve Bayes," in *Proc. Int. Conf. Appl. Techn. Cyber Secur., Intell. Appl. Techn. Cyber Secur. Intell. (ATCI)*. Springer, 2019, pp. 281–298, doi: 10.1007/978-3-319-98776-7_30.

[19] F. Iqbal, J. M. Hashmi, B. C. M. Fung, R. Batool, A. M. Khattak, S. Aleem, and P. C. K. Hung, "A hybrid framework for sentiment analysis using genetic algorithm based feature reduction," *IEEE Access*, vol. 7, pp. 14637–14652, 2019, doi: 10.1109/ACCESS.2019.2892852.

[20] Y. M. Wazery, H. S. Mohammed, and E. H. Houssein, "Twitter sentiment analysis using deep neural network," in *Proc. 14th Int. Comput. Eng. Conf. (ICENCO)*, Dec. 2018, pp. 177–182, doi: 10.1109/ICENCO.2018.8636119.

[21] S. Sharma and A. Jain, "An empirical evaluation of correlation based feature selection for tweet sentiment classification," in *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, 2020, pp. 199–208, doi: 10.1007/978-981-15-3125-5_22.

[22] J. Majidpour and K. Al-Barznji, "Opinions for receiving COVID-19 vaccines based on sentiment analysis," *J. Pharmaceutical Negative Results*, vol. 13, no. S01, 2022, doi: 10.47750/pnr.2022.13.s01.197.

[23] U. A. Haje, M. H. Abdalla, R. M. S. Kurda, and Z. M. Khalid, ‘‘A new model for emotions analysis in social network text using ensemble learning and deep learning,’’ *Academic J. Nawroz Univ.*, vol. 11, no. 1, pp. 130–140, Mar. 2022, doi: 10.25007/ajnu.v11n1a1250.

[24] R. A. Khurma, I. Aljarah, A. Sharieh, M. A. Elaziz, R. Damaševičius, and T. Krilavičius, ‘‘A review of the modification strategies of the nature inspired algorithms for feature selection problem,’’ *Mathematics*, vol. 10, no. 3, p. 464, Jan. 2022, doi: 10.3390/math10030464.

[25] A. Zarshenas and K. Suzuki, ‘‘Binary coordinate ascent: An efficient optimization technique for feature subset selection for machine learning,’’ *Knowl.-Based Syst.*, vol. 110, pp. 191–201, Oct. 2016, doi: 10.1016/j.knosys.2016.07.026.

[26] D. Saltiel and E. Benhamou, ‘‘Feature selection with optimal coordinate ascent (OCA),’’ 2018, *arXiv:1811.12064*. Accessed: Aug. 16, 2023.

[27] D. Saltiel, E. Benhamou, R. Laraki, and J. Atif, ‘‘Trade selection with supervised learning and optimal coordinate ascent (OCA),’’ in *Proc. 5th ECML PKDD Workshop Mining Data Financial Appl. (MIDAS)*. Ghent, Belgium: Springer, Sep. 2020, pp. 1–15, doi: 10.1007/978-3-030-66981-2_1.

[28] C. Zuo, J. Qian, S. Feng, W. Yin, Y. Li, P. Fan, J. Han, K. Qian, and Q. Chen, ‘‘Deep learning in optical metrology: A review,’’ *Light, Sci. Appl.*, vol. 11, no. 1, p. 39, Feb. 2022, doi: 10.1038/s41377-022-00714-x.

[29] S. Bhattacharya, S. R. K. Somayaji, T. R. Gadekallu, M. Alazab, and P. K. R. Maddikunta, ‘‘A review on deep learning for future smart cities,’’ *Internet Technol. Lett.*, vol. 5, no. 1, Jan. 2022, Art. no. e187, doi: 10.1002/itl2.187.

[30] J. Yu, A. de Antonio, and E. Villalba-Mora, ‘‘Deep learning (CNN, RNN) applications for smart homes: A systematic review,’’ *Computers*, vol. 11, no. 2, p. 26, Feb. 2022, doi: 10.3390/computers11020026.

[31] P. Patel, D. Patel, and C. Naik, ‘‘Sentiment analysis on movie review using deep learning RNN method,’’ in *Intelligent Data Engineering and Analytics: Frontiers in Intelligent Computing: Theory and Applications*, vol. 2. Singapore: Springer, 2020, doi: 10.1007/978-981-15-5679-1_15.

[32] M. A. Aziz, M. M. Islam, M. S. P. Munshi, S. Samshi, and M. M. Khan, ‘‘Sentimental analysis on reviews of protein supplement using RNN-LSTM,’’ in *Proc. 6th Int. Conf. Comput. Methodol. Commun. (ICCMC)*, Mar. 2022, pp. 796–800, doi: 10.1109/ICCMC53470.2022.9753901.

[33] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, ‘‘Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions,’’ *J. Big Data*, vol. 8, no. 1, pp. 1–74, Mar. 2021, doi: 10.1186/s40537-021-00444-8.

[34] J. A. Nasir, O. S. Khan, and I. Varlamis, ‘‘Fake news detection: A hybrid CNN-RNN based deep learning approach,’’ *Int. J. Inf. Manage. Data Insights*, vol. 1, no. 1, Apr. 2021, Art. no. 100007, doi: 10.1016/j.jjimei.2020.100007.

[35] D. P. Kingma and J. Ba, ‘‘Adam: A method for stochastic optimization,’’ *arXiv:1412.6980*. Accessed: Aug. 16, 2023.

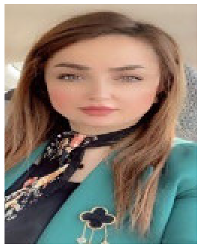
[36] J. Han and C. Moraga, ‘‘The influence of the sigmoid function parameters on the speed of backpropagation learning,’’ in *Proc. Int. Workshop Artif. Neural Netw.*, in Lecture Notes in Computer Science, 1995, pp. 195–201, doi: 10.1007/3-540-59497-3_175.

[37] D. Scherer, A. Müller, and S. Behnke, ‘‘Evaluation of pooling operations in convolutional architectures for object recognition,’’ in *Proc. Int. Conf. Artif. Neural Netw.* Berlin, Germany: Springer, 2010, doi: 10.1007/978-3-642-15825-4_10.

[38] D. M. Allen, ‘‘The relationship between variable selection and data agumentation and a method for prediction,’’ *Technometrics*, vol. 16, no. 1, pp. 125–127, Feb. 1974, doi: 10.1080/00401706.1974.10489157.

MOHAMMED HUSSEIN ABDALLA was born in Sulaymaniyah, Iraq, in 1989. He received the B.Sc. degree in computer science from the University of Sulaimani, Sulaymaniyah, in 2012, and the M.Sc. degree in software engineering from Fırat University, Elâz?ğ, Turkey. In 2013, he joined the University of Raparin, as an Assistant Programmer, where he became an Assistant Lecturer, in 2022. His current research interests include deep learning, machine learning, the IoT, and NLP.

JAFAR MAJIDPOUR received the master's degree in software engineering from the East Azarbaijan Science Research Branch, Tabriz, Iran, in 2013. Since 2013, he has been with the Department of Basic Education, University of Raparin, Iraq, as a Lecturer. His primary research interests include computer vision, machine learning, deep learning, and image processing.



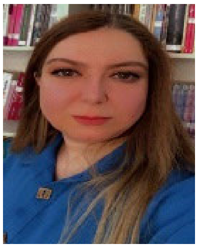
RAHELA AZIZ RASUL was born in Sulaymaniyah, Iraq, in 1988. She received the B.S. degree in statistics from the University of Sulaimani, Sulaymaniyah, in 2011, and the M.Sc. degree in statistics from Van Yüzüncü Yıl University, Van, Turkey. In 2013, she joined the University of Raparin, as an Assistant Programmer, where she became an Assistant Lecturer, in 2022. Her current research interests include statistics, linear regression, and feature selection methods.

ABDULRAHMAN A. ALSEWARI (Senior Member, IEEE) was born in Sana'a, Yemen, in 1980. He received the B.Eng. degree in computer engineering from the Military College of Engineering, Baghdad, Iraq, in 2002, and the M.Sc. degree in electronic system design engineering and the Ph.D. degree in software engineering from Universiti Sains Malaysia, in 2008 and 2012, respectively. From 2013 to 2022, he was an Associate Professor and a Researcher with the Department of Software Engineering, Faculty of Computing, Universiti Malaysia Pahang. He is currently the Machine Learning Laboratory Leader, a Seminar Coordinator, and a Senior Lecturer with the College of Computing, Faculty of Computing, Engineering, and the Built Environment, Birmingham City University, U.K. He has authored or coauthored one book, more than 80 articles, and more than ten inventions. His research interests include software engineering, software testing, optimization algorithms, artificial intelligence, soft computing, embedded systems, and image processing.

TARIK AHMED RASHID (Member, IEEE) received the Ph.D. degree in computer science and informatics from the College of Engineering, Mathematical and Physical Sciences, University College Dublin (UCD), in 2006. He was a Postdoctoral Fellow with the School of Computer Science and Informatics, UCD, from 2006 to 2007. In 2017, he joined the University of Kurdistan Hewlêr. His research interests include three fields. The first field is the expansion of machine learning and data mining to deal with time series applications. The second field is the development of DNA computing, optimization, swarm intelligence, and nature-inspired algorithms and their applications. The third field is networking, telecommunication, and telemedicine applications. He is a member of ACM.

ARAM MAHMOOD AHMED received the B.Sc. degree in software engineering from Koya University, Iraq, in 2011, the master's degree in computer network technology from Northumbria University, U.K., in 2013, and the Ph.D. degree in optimization and metaheuristic algorithms from Sulaimani Polytechnic University, Iraq, in 2021. He is currently the Head of the Information Technology Department, Kurdistan Institution for Strategic Studies and Scientific Research, Sulaymaniyah, Iraq. He is interested in modeling biological and natural systems into computational techniques. He is also an active researcher who has published books and papers in peer-reviewed academic journals with high impact factors.

BRYAR A. HASSAN received the M.Sc. degree in software engineering and the Ph.D. degree in computer science from the University of Southampton, U.K., in 2013 and 2021, respectively. He is currently an Assistant Professor with the Department of Computer Science, College of Science, Charmo University. His research interests include meta-heuristic algorithms, data mining, clustering, swarm intelligence, artificial intelligence, and nature-inspired algorithms.



NOOR BAHJAT TAYFOR received the bachelor's degree in software engineering from the University of Mosul, in 2007, and the master's degree in advanced computer science from The University of Liverpool, in 2015. In October 2016, she started the academic career. She is currently an Assistant Lecturer with the Information Technology Department, Kurdistan Technical Institute. Her research interests include artificial intelligence, natural language processing, pattern recognition, machine learning, text mining, and sentiment classification.

SHKO MUHAMMED QADER received the bachelor's degree in software engineering from the College of Engineering, Salahaddin University-Erbil, in July 2007, and the M.Sc. degree in software engineering with management studies from the College of Engineering, Faculty of Science, Engineering and Computing, Kingston University London, in 2013. In 2015, he joined the Computer Science Institute, Slemani Polytechnic University. His research interests include artificial intelligence, machine learning, data mining, optimization algorithms, medical computing, semantic web, and NLP. He was selected for fellowship programs at both CDAC, India, and KOIKA, South Korea.

SINAN Q. SALIH received the B.Sc. degree in information systems from the University of Anbar, in 2010, the M.Sc. degree in computer science from the College of Information Technology, Universiti Tenaga Nasional (UNITEN), in 2012, and the Ph.D. degree in soft computing and intelligent systems from the Faculty of Software Engineering, University of Malaysia Pahang (UMP), in 2019. He is currently a Lecturer with the Technical College of Engineering, Al-Bayan University. He has published more than 70 research papers, and coauthored with more than 50 researchers. His current research interests include optimization algorithms, nature-inspired metaheuristics, machine learning, and feature selection problems for real-world problems.