Proceedings of the Jangjeon Mathematical Society
28 (2025), No. 4, pp. 657 - 674

www.jangjeon.or.kr
http://dx.doi.org/10.17777/pjms2025.28.4.019

HYBRID DEEP LEARNING-BASED MODELS FOR SOCIAL MEDIA SENTIMENT ANALYSIS

ANJANI KUMAR SINGHA, SHAMS SIDDQUI,
PRADEEP KUMAR TIWARI, *GAURAV YADAV
AND [5]SURYA KANT

ABSTRACT. There have been numerous applications for expression evaluation of public assertions made on social media platforms like Twitter and Facebook; however, many obstacles exist to overcome. With increasingly complicated learning algorithms, hybrid approaches have proven to be effective models for lowering impression errors. The main aim to enhance the accuracy and efficiency of sentiment classification by integrating multiple deep learning architectures, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and Transformer-based models (e.g., BERT, GPT). Our research paper emphasizesComparing the accuracy, precision, recall, and F1-score of hybrid models against standalone deep learning approaches (e.g., LSTM, CNN, BERT). Eight linguistic retweets and evaluation samples by a wide range of disciplines to build and test mixed-depth sentiment classification learning approaches that integrate support vector machines (SVM), convolutional neural networks (CNN), and long short-term memory (LSTM). SVM, LSTM, and CNN are three single models against which the two technologies are contrasted. In assessing every method, calculation time and accuracy were taken into account. Across all varieties of datasets, hybrid models improved sentiment classification accuracy in comparison to base classifiers, notably when deep learning algorithms and SVMs were combined, and python experiment and Google Cloab and dataset took by Sentiment140 (10%) 160.000, Tweets Airline 4.726, IMDB movie reviews (1) 50.000 and IMDB movie reviews (2) 25.000

KEYWORDS:Support vector machines (SVM), convolutional neural networks (CNN), and long short-term memory (LSTM).

Gaurav Yadav: Gyadav444@gmail.com, Postal Address: 17/238 Z-27/III Baldev Vihar Colony Sasni Gate Aligarh-202001, UP , India

Submission Date: 10.10.2024

1. INTRODUCTION

Nowadays, the researcher on data sentiment analysis through social media platforms like Twitter and Facebook are gaining popularity. In the sentiment analysis, also known as opinion mining, plays a crucial role in understanding public sentiment across various domains, including politics, marketing, and social behaviour. Even though extensive work has already been done in this field, there are numerous problems that need fixing, like the rising simulation dependability, speeding up execution, and using methods that were designed specifically for specific kinds of information and data domains [1]. The machine is clear that learning models have been widely used while discussing sentiment analysis recently, and their enormous capability has been discovered. Many studies, such as those on marketing tactics, financial planning, and healthcare analysis, are solely concerned with creating a single component from a single dataset(s) in the relevant area. Sentiment polarity-based machine learning algorithms used to tweet be extensively discussed for social media network usage. CNN models and recurrent neural networks (RNN) were shown by Mohamed and Hussain [2] to be capable of overcoming the limitation of brief models that use text for deep learning. In addition, the research of Qian et al. [3] revealed the efficacy of LSTM on Twitter, climate, and emotion at various text levels. After reviewing several recent studies, we discovered that RNN and CNN outperform techniques with comparatively high overall efficiency—any value basic neural network models and CNN.

In contrast to shallow neural network models, deep neural networks benefit from extracting features while training from large datasets. The main reason for this is that deep models can remove and develop better characteristics than shallow models, and they do this by using intermediary nested loops. Deep neural networks have the potential to be significantly more computationally and parameter-efficient while maintaining the same level of precision. Considerable emphasis is placed on this; deeper neural networks develop various, increasingly abstract data representations at each layer. Every deep learning strategy has benefits and drawbacks, even if a specific machine learning technique can be successfully applied in several areas. CNN needs fewer parameters and less supervision than LSTM, but processing takes longer and typically produces worse outcomes. The LSTM, on the other hand, takes longer to comprehend but improves performance for a long prison sentence. The concept of merging multiple (or more) procedures is presented as a way to take advantage of both their benefits and address some of the drawbacks of existing approaches. Alfrjani et al. [4] coupled machine learning and linguistic knowledge, improving sentiment analysis's precision on evaluations.

Another example is the hybrid approach. Gupta and Joshi [5] suggested using Twitter's sentiment analysis. It mixes lexicon and machine learning. Hence, the best solution is a hybrid system with built-in approved material equipped to handle any potential drawbacks linked to a single system, if any exist. Depending on the task, the combined models' efficacy may change. Machine learning lexical analysis CNN with RNN and e-CNN supplemented by SVM enhanced outcomes throughout the board. When analysing the mood of various areas and kinds of datasets, the mixture of CNN, LSTM, and SVM intends to benefit from the two deep network structure models and SVM algorithms. Also, other sorts of input information are gleaned through social media platforms, including ratings and comments. The data required also incorporates distinctions within and between different kinds, such as the proportion of tweeter and comment durations, the variation of themes in every data, the response rate, and the degree to which expressed emotions and extraneous details are present. Several techniques could have been more accurate or performed better in sentiment analysis across diverse domains. This researcher proposes a Hybrid Deep Learning-based Model that leverages the strengths of multiple deep learning architectures to enhance sentiment classification accuracy.

As a result, some methodologies could not work well or be challenging to use with particular kinds of input information. Our study raises the question without regard to the dataset's characteristics. As a result, our research investigates how particular hybrid models respond to various datasets from multiple areas. In this paper, the three models, LSTM, SVM, and CNN, were combined and assessed. Many existing hybrid deep learning models perform well on specific datasets (e.g., Twitter, Reddit) but struggle to generalize across platforms due to differences in linguistic styles, slang, and context. Social media sentiment analysis faces challenges such as data sparsity, context ambiguity, sarcasm detection, and multimodal inputs (text, images, emojis).

1.1. CONTRIBUTION OF PAPER

- Clarify whether techniques from image processing (e.g., CNNs for feature extraction) have been adapted for text-based sentiment analysis.
- justify their inclusion by highlighting shared challenges in sentiment classification across languages (e.g., noisy data, dialectal variations).

.

2. RELATED WORK

The project aims to create a hybrid approach to the sentiment analysis method to increase accuracy. The approaches suggested and utilized in other studies that have initially been looked at are detailed below. Hybrid models can be created in various ways—the researchers integrated an SVM and CNN model to improve the precision of image processing. SVM serves as a recognition system, while features are extracted using a convolutional network layer. Softmax functions are utilized with the original CNN. For the textual categorization of negative and positive attitudes, Srinidhi et al. [6] suggested a definition like a model that combines elements of both incorporated SVM and LSTM with such a radial foundation function kernel. A mixed model was assessed on the IMDB book review datasets. These models include many deep learning techniques for identification using SVM. Several of them are used in picture-based face recognition. Our study performs classification using SVM or ReLU after combining two deep learning algorithms.

A blended deep-learning framework developed by K.Paulraj et al. [7] is very effective for sentiment analysis in languages with limited resources. They employed CNN to learn sentiment embedding vectors and SVM to classify sentiment. Four samples in Hindi representing different topics were used to test the model. A multilingual LSTM-CNN model was used by Vo et al. [8] to analyze sentiment in comments and feedback from e-commerce websites. Rehman et al. [9] also use hybrid CNN-LSTM models to study the sentiment in film reviews. The same methods are applied in numerous publications, for instance. A technique known as a hybrid heterogeneous support vector machine was created by Kaur et al. (H-SVM). They conducted sentiment classification on COVID-19-related Twitter data. Three deep learning models, including CNN, LSTM, and CNN-LSTM, were used by Kastrati et al. [10] to categorize Facebook comments about the COVID-19 epidemic.

They used the contextualized word immersion model BERT and a classifier word embedding technique dubbed FastText to understand and create vector representation. Both studies rated Tweets and comments as favorable, negative, or neutral.

Such approaches were only tested on a small number of sample datasets or diverse datasets from a single domain. Hence, their authenticity has only sometimes been established by Offering a generalized model for sentiment analysis; Jnoub et al.'s [11] study focuses on fusing CNN with a unique technique to convert reviews to vectors. IMDb, reviews, and a personal database compiled from customer reviews were used to assess the program. A hybrid deep learning model that includes LTSM and CNN was proposed by Ombabi et al. [12]. The Arabic language also uses FastText for word embedding and SVM for identification. For embedding, our research utilized both BERT and Word2Vec both used.

For identifying combined retweets and comments, we suggested four different Algorithms for hybrid deep learning: SVM, LSTM, and CNN. Moreover, additional studies integrate emotion words and polarity-changing devices with Analysis based on machine learning and a lexicon. The "customer and material sentiment classification" issue is addressed in Sanchez-Rada and Iglesias' study [13]. A hybrid model that combines characteristics from many social context levels was proposed by %ey. The model is assessed using various datasets. In a study by Samih Amina et al. [14], a hybrid strategy was described in which a preliminary recommendation list produced by combining data mining techniques and content-based algorithms was improved by using sentiment analysis of movie evaluations. Wang Q et al. [15] suggested using a sentiment classifier generated as a secondary filter following content-based filtering from movie reviews using the same methodology. These studies employ conventional methods for sentiment analysis. Our research uses deep learning methods to increase sentiment categorization accuracy. Recognition systems have already been successfully used in sentiment analysis, where lower bandwidth layers are trained on elevated controlled datasets like XLNET [16] and BERT [17], which were proposed by researchers at Google AI.

3. TECHNIQUE

This research assesses four hybrid models in light of all the benefits and potential of these models to enhance sentiment analysis methods. The three primary focuses of the technique are the data to be used, the procedure for creating feature vectors, and the development of Hybrid techniques for an appropriate answer to sentiment analysis.

3.1. DATASET

Our study is concentrated on evaluating generic application models rather than fixing an issue in a specific field. Rather than creating and categorizing fresh datasets for a particular application domain for this investigation, we used several publicly available datasets. The choosing process considered several factors, including the ability to protect personal information acceptability in educational institutions, variety of resources and themes, and size. The chosen datasets allow for a thorough evaluation of the sentiment analysis techniques explored in this work. The research aims to determine if the models consistently deliver reliable data regardless of the kind and size of the dataset.

Eight datasets were used in the studies. Five datasets feature evaluations, whereas three (Sentiment140, Tweets Airline, and Tweets SemEval) contain tweets.The largest of the three Twitter datasets, Sentiment140, contains 1.6 million tweets with labels for negative or positive sentiment. At the same time, Tweets SemEval and Tweets Airline each have 15,640 and 18,650 tweets with neutral, negative, or positive labels. 125,000 remarks from user ratings of literature, audio, and movies—labeled as

either positive or negative—are included in the five review datasets (IMDB movie reviews, IMDB film reviews, Cornell movie reviews, and IMDB book and music reviews. Further information on them is provided in [1]. After reviewing the acquired data, we discovered that seven of the eight data sources have starting labels of positive and negative, with a roughly equal proportion of each Airline and Twitter SemEval data source, including neutrality labeling and positive and negative labels. When learning models and performing classification, it is crucial to have a neutral classification performance to prevent posterior probabilities from being skewed. We focus on either positive or negative polarity sentiment analysis in this study. The neutral labels were eliminated to decrease the size of these databases. The balance between the remaining negative and positive classes is evaluated. We also used a k-fold bridge on the data to assess the models.In this manner, bias against a certain portion of the information is avoided, and the tests cover all occurrences of the datasets. Table 1 shows the number of samples (both positive and negative) taken from each database for investigations.

3.2. DATA PREPROCESSING AND CHARACTER VECTOR CREATION

The content, the statement, and the component or feature can all be used as stages of separation for sentiment categorization. In our trials, we used word embedding methods and document-based sentiment classification on eight databases of tweets and reviews. Text-training data must be cleaned before being used as a source for classification methods in sentiment analysis. White space, punctuation, and line breaks are all examples of unnecessary text or phrase data that are deleted. TF-IDF and embedding are two approaches that are frequently used for this task. Since it produces superior results to TF-IDF, our suggestion uses the latter [1]. The feature vector was then constructed using BERT and Word2vec word embedding models in Table 1.

TABLE 1 LISTS THE NUMBER OF DATASET SAMPLES.

Number of samples	#	Datasets
20.000	IMDb movie reviews (2) 5	IMDB movie reviews (1)
11.662	Cornell movie reviews6	Tweets SemEval
2.000	Book reviews7	Tweets Airline
2.000	Music reviews8	Sentiment140 (10%)

Developers at Google AI Language released the BERT language model for natural language processing [17]. Following Word2vec in development, BERT has some improvements over Word2vec, including support for out-of-vocabulary (OOV) terms.

Tomas Mikolov released Word2Vec in 2013 at Google. The training data for this uncontrolled learning method came from a sizable corpus[18]. With a matrix NxD, where N is the number of articles, and D is the degree of word embedding, Word2vec has a substantially smaller size beyond one encoding. The skip-gram and continuous bag-of-words (CBOW) models are included in Word2vec[19]. Such theories are based on the likelihood that words will appear close to one another[20]. We can begin with such a syllable and use skip-gram to forecast the words that will likely follow it. Unfortunately, one of the main problems with Word2Vec is that it needs to handle terms that aren't commonly used[21]. We use the unique token [UNK] for terms not in the dictionary to avoid this issue.

Additionally, we reduce the usage of the special token by retraining the Word2vec model using our vocabulary datasets and any terms that occur at least six times[22]. The varied lengths of the database samples present a challenge when modeling sentiment, whereas constant input vectors are necessary for deep learning models [23]. Since being cleaned, the collections of ratings and comments are shown as graphs in Figures 1 and 2. The size of the dataset is shown on the x-axis, and the rate of occurrence is shown on the y-axis. Since we selected several datasets from various sources, many graphs are pretty jagged [24]. It may be possible to fit the models by standardizing the data and defining smooth borders according to the sample size.

3.2.1. TEXT CLEANING

- Removal of special characters, punctuation, URLs, mentions (@username), and hashtags.
- Conversion of all text to lowercase to maintain consistency.

3.2.2. TOKENIZATION

- Splitting text into individual words or subwords using word-based or subword-based tokenization (e.g., WordPiece, Byte Pair Encoding).

3.2.3. STOP WORD REMOVAL

- Elimination of commonly used words (e.g., "the," "is," "in") that do not contribute to sentiment detection.

3.3. HYBRID APPROACHES SEVERAL TECHNIQUES

A hybrid model for sentiment analysis. In this study, we experimented with combining various effective strategies. As seen in Figure 1, we first generate the feature vector using Word2vec or a retrained BERT model. The order of the LSTM and CNN models used in the subsequent stages is then changed to either

(1)Word2vec/BERT->CNN->LSTM

(2)Word2vec/BERT->LSTM->CNN.

We also change the model's final phase, employing an SVM or a ReLU activation function.

During our tests, feature vectors were created using two methods [25]. Word2vec was initialized with respect to the weights to serve as the initial method for learning the integration of all phrases in the learning set of data. We used BERT as our second strategy because Word2vec does not include situational Analysis to deal with complex semantic and syntactic or polymorphic instances in natural languages. Inside this investigation, a pertained BERT model was used [26]. The BERT model was utilized as a feature representation to produce raw data for such suggestions of different models once the variables were adjusted. The comment and rating data were loaded into the BERT model to create the feature vectors that serve as the source for the other models that carry out the categorization.

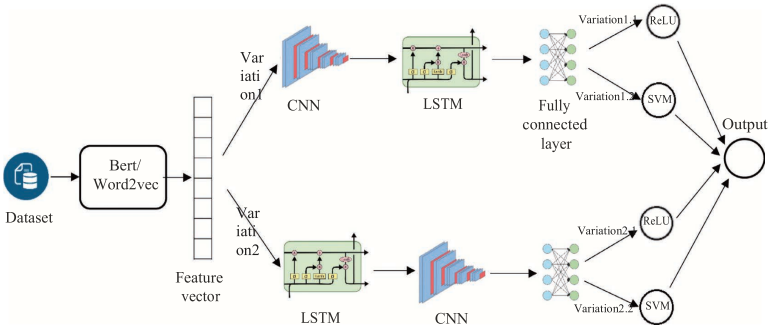

FIGURE 1: METHODOLOGY FOR CONDUCTING SENTIMENT ANALYSIS.

The following phase includes LSTM and CNN deep learning models, which are employed due to their successful sentiment assessment and efficiency. They also utilize the two network structures while doing sentiment analysis of data from various domains. Given that it consists of numerous layers that analyze and transfer data in one direction, from beginning to end, without cycles, a CNN is a sort of feed-forward neural network. It utilizes a deep neural network design with the usual starting point of convolution and pooling/subsampling layers that alter inputs before feeding into a densely integrated classification stage. A single convolutional neural network (1D CNN) was employed in this study. Among the various variants of the RNN design is the LSTM. An LSTM block comprises the output and input sections, the storage unit, and three so-called gates—the forgetting gate, input value, and output unit. While RNNs are good at handling temporal indicators, CNNs excel at handling vertically relevant data. Multilayer CNN and LSTM can learn and remember both series data and specific details. Hence, the combined system utilizes the best aspects of the two separate temporal and spatial worlds.

3.4. RELEVANCE TO SOCIAL MEDIA SENTIMENT ANALYSIS

Explain why the chosen dataset is well-suited for your research, such as its real-world applicability, diverse linguistic styles, or volume of labelled data.

4. PROPOSED HYBRID MODEL

Throughout this paper's scope, we present four different hybrid deep learning models based on different ways of using LSTM and CNN in deep learning layers, and different ways of using CNN and SVM in classification layers are presented. Tables 2 and 3 display the hybrid models' architectural details, and the following section goes into more detail.

4.1. SCENARIO COMBINATION 1.

In the first hybrid model, CNN and LSTM models are combined. Table 2 lists the visualization of the model connection, the connection procedure, and the information processing workflow. The hidden layer known as "%e function encapsulation," which seems initialized using random weights, can discover the embedding for every word inside the training sets. The CNN, the first level of the hybrid model, is given the vector created by embeddings. It has several convolution operations with kernel sizes of 3, each containing 512, 256, and 128 filters that take in and analyse data while supplying it to the following deep learning layer.

In the CNN-LSTM is computationally faster and more robust to noise, making it ideal for real-time applications.

TABLE 2: A HYBRID CNN-LSTM MODEL.

Param #	Layer (type)	Output shape
129	dense_3 (dense)	(None, 1)
16,512	dense_2 (dense)	(None, 128)
622,720	flatten_1(flatten)	(None, 128)
0	conv1d_6(Conv1D)	(None, 500)
98,432	conv1d_5 (Conv1D)	(None, 38, 128)
393,472	conv1d_3 (Conv1D)	(None, 38, 256)
568,512	lstm_2 (LSTM)	(None, 38, 512)
1,205,100	embedding_1(embedding)	(None, 38, 100)

Nontrainable params: 2,108,720
Total params: 4,483,675
Trainable params: 2,785,386

In the LSTM-CNN performs better on long-form text but requires more computational resources due to early sequential learning.

Table 3: A hybrid LSTM-CNN model.

Param #	Layer (type)	Output shape
129	dense_6 (dense)	(None, 1)
16,512	dense_5 (dense)	(None, 128)
622,720	dense_4 (dense)	(None, 128)
0	flatten_1(flatten)	(None, 500)
98,432	conv1d_6(Conv1D)	(None, 38, 128)
393,472	conv1d_5 (Conv1D)	(None, 38, 256)
768,512	conv1d_3 (Conv1D)	(None, 38, 512)
1,202,000	lstm_2 (LSTM)	(None, 38, 100)
1,808,900		
embedding_2(embedding)		
Total params: 5,910,687		
Trainable params: 4,108,787		
Nontrainable params: 2,608,700		

The LSTM, the hybrid model's second layer, generates a 1500 matrix passed to the classification. The hybrid model predictor comprises two uninterrupted, completely interconnected layers with 128 nodes, multiple output layers, and an activation function.

4.2. HYBRID DEEP LEARNING MODEL ARCHITECTURE

Our hybrid model combines Support Vector Machine (SVM) for feature-based classification, Convolutional Neural Networks (CNNs) for feature extraction, and Long Short-Term Memory (LSTM) networks for capturing sequential dependencies in sentiment analysis. The detailed architecture is as follows:

4.2.1. CONVOLUTIONAL NEURAL NETWORK (CNN) FOR FEATURE EXTRACTION

- Embedding Layer: Pre-trained word embeddings (GloVe/Word2Vec) with a dimension size of 300.
- Convolutional Layer: 1D CNN with 64 filters, a kernel size of 5, and ReLU activation to detect local sentiment patterns.

4.2.2. LONG SHORT-TERM MEMORY (LSTM) FOR SEQUENTIAL DEPENDENCY LEARNING

- LSTM Layer: 128 hidden units to capture long-range dependencies. Dropout (0.3) to prevent overfitting.

4.2.3. FULLY CONNECTED (DENSE) LAYER FOR CLASSIFICATION

- Flatten Layer: Converts feature maps into a one-dimensional vector. Dense Layer (Intermediate):
  - 128 neurons with ReLU activation.
  - Dropout (0.3) for regularization.
- Output Layer:
  - Softmax activation for multi-class sentiment classification (positive, negative, neutral).

4.2.4. SVM FOR FINAL CLASSIFICATION (HYBRID APPROACH)

- Regularization parameter (C): 1.0 for balanced complexity and accuracy.

5. RESEARCH FINDING

That section covers the tests done to evaluate the effectiveness of the suggested hybrid models. In addition, we investigate additional prevalent deep learning models (SVM, CNN, and LSTM). The pre-trained word embeddings (GloVe, Word2Vec, or Transformer-based embeddings) . All of these were evaluated using the eight samples presented in subsection 3.1 that had undergone character recognition pre-processing. Reliability, AUC, and F-score were used to assess the models' effectiveness across all trials. We also display memory and accuracy because those are the two metrics from which the F-score is built. Sections 5.2 and 5.3 present, examine, and comment on the findings.

5.1. PERFORMANCE EVALUATION

The essential hardware components, library services, and relevant variable setup were completed before the investigations. With the GPU Tesla P100-PCIE16GB or GPU Tesla V100-SXM2-16GB, the Keras and TensorFlow libraries and Google Colab Pro were used. In all trials, we set the parameters for our code to echoes 4, k-fold 10, batch size 32 for ratings, and batch size 128 for tweeting. The K-fold validation method has three typical choices: k=4, k=6, and k=11. When applying machine learning to assess models, K=10 is the most widely used value. The latter value is used whenever the dataset is large enough for the subgroups to have many instances. The case is present in the datasets used in this investigation. For each of the 10 validations, one part acts as the testing set and nine parts serve as the training set to be sure that every train or sample solution is sizable enough to reflect the dataset and that a value of k is selected. Importantly, this process guarantees that the k setups in all verifications have the same size as the k models in the cross-validation, which are both generated from training sets that are identical in size. It is advised to divide the data into similar groups to compare the predictive accuracy.

5.2. RESULTS

The results of eight experiments in each set are displayed, with three baseline models (SVM, CNN, and LSTM) and four hybrid models (CNN-LSTM-SVM, CNN-LSTM-CNN, and LSTM-CNN-SVM, respectively), as well as three baseline models (SVM, CNN, and LSTM). In a comparative analysis that is also included, the outcomes of the suggested hybrid methods are compared to those of the

previous methods. Our tests were carried out twice: using Word2vec to train word embedding and once using a pertained BERT model. Tables 4–8 explain the trial outcomes using Word2vec and BERT, since BERT consistently produced good outcomes. The comparison findings between Word2vec and BERT are shown using side-by-side bar charts in Figures 4–8. When a pertained BERT model is used to retrieve a feature vector, the calculated values for all samples and classification models are very good (around 90%), with Tweets Airline and IMDB movie reviews showing particularly high accuracy (92.9% and 93.4%, respectively) (1). Also, the results show that hybrid models (SVM, CNN, or LSTM) are more accurate than standalone deep learning models (SVM, CNN, or LSTM) for seven out of eight datasets. The reliability scores for CNN for using Word2vec in the music evaluation and book review databases are 76.4% and 76.5%, respectively. When the LCNN-SVM model is used, the results increase dramatically to 83.7% and 82.7%, improving 7.3% and 6.2%, respectively.

In seven out of eight datasets for the F-score (Table 7), hybrid models created greater (or equal) results than pure deep learning models. Compared to the single deep learning model, the hybrid models outperform the AUC value in Table 8; in six of eight datasets employing Word2vec, hybrid models using SVM for categorization produced the best performance. The datasets with the best ratings for all metrics in all scenarios are the Twitter Airline Database and IMDB Movie Ratings (1). With hybrid LSTMCNN and LCNN-SVM algorithms, ratings of books and music perform well. All algorithms' reliability for the %e Sentiment140 dataset is low. Figures 1 and 2 show how the average number of samples in the dataset varies with the duration of sample data. Furthermore, unique compared to the other datasets is the "%e Sentiment140" dataset. The length data samples have many different measurement samples in show figure 2,3,4,5, and 6.

5.3. DISCUSSION

As shown in Figure 2, pre-trained BERT outperforms Word2vec for sentiment analysis across all models and datasets.

TABLE 4 ACCURACY OF COMPARES CORRECTNESS FOR VARIOUS DATASETS.

Datasets	LC-S	CL-S	L-C	C-L	LTSM	CNN	SVM	LC-S	CL-S	L-C	C-L
LTSM- CNN											
Music reviews	83.9	83.5	84.1	84.0	84.1	84.0	82.4	79.9	80.6	79.9	80.1
80.3	79.7										
Book reviews	92.7	92.9	92.8	92.8	92.8	92.9	92.0	87.8	88.6	87.5	88.0
87.1	86.8										
Cornell movie reviews	91.8	91.7	91.8	91.9	91.8	91.9	91.2	85.7	85.6	86.2	85.0
86.4	84.4										
IMDb movie reviews (2)	93.4	93.4	93.4	93.4	93.4	93.4	93.3	90.3	90.0	90.0	89.7
88.5	87.6										
IMDb movie reviews (1)	90.7	90.7	90.7	90.7	90.6	90.7	90.5	89.4	98.4	89.4	89.2
85.1	87.3										
Tweets SemEval	87.1	86.9	87.0	86.9	86.9	87.0	85.3	75.9	76.2	76.4	73.0
76.1	72.4										
Tweets Airline	91.0	90.4	91.1	90.7	91.0	90.7	89.9	83.7	78.7	83.5	75.8
77.2	76.2										
Sentiment140 (10%)	89.2	89.5	89.1	89.0	89.2	89.2	87.8	82.7	76.6	82.1	70.9

79.6 76.5

TABLE 5 RECALL COMPARES THE CORRECTNESS OF VARIOUS DATASETS.

Datasets	LC-S	CL-S	L-C	C-L	LTSM	CNN	SVM	LC-S	CL-S	L-C	C-L	LTSM	CNN
Music reviews	83.9	83.5	84.1	84.0	84.1	84.0	82.4	79.9	80.6	79.9	80.1	80.3	79.7
Book reviews	92.7	92.9	92.8	92.8	92.8	92.9	92.0	87.8	88.6	87.5	88.0	87.1	86.8
Cornell movie reviews	91.8	91.7	91.8	91.9	91.8	91.9	91.2	85.7	85.6	86.2	85.0	86.4	84.4
IMDb movie reviews (2)	93.4	93.4	93.4	93.4	93.4	93.4	93.3	90.3	90.0	90.0	89.7	88.5	87.6
IMDb movie reviews (1)	90.7	90.7	90.7	90.7	90.6	90.7	90.5	89.4	98.4	89.4	89.2	85.1	87.3
Tweets SemEval	87.1	86.9	87.0	86.9	86.9	87.0	85.3	75.9	76.2	76.4	73.0	76.1	72.4
Tweets Airline	91.0	90.4	91.1	90.7	91.0	90.7	89.9	83.7	78.7	83.5	75.8	77.2	76.2
Sentiment140 (10%)	89.2	89.5	89.1	89.0	89.2	89.2	87.8	82.7	76.6	82.1	70.9	79.6	76.5

.

TABLE 6 COMPARES F-SCORES FOR VARIOUS DATABASE TYPES

Datasets	LC-S	CL-S	L-C	C-L	LTSM	CNN	SVM	LC-S	CL-S	L-C	C-L	LTSM	CNN
Music reviews	83.9	83.5	84.1	84.0	84.1	84.0	82.4	79.9	80.6	79.9	80.1	80.3	79.7
Book reviews	92.7	92.9	92.8	92.8	92.8	92.9	92.0	87.8	88.6	87.5	88.0	87.1	86.8
Cornell movie reviews	91.8	91.7	91.8	91.9	91.8	91.9	91.2	85.7	85.6	86.2	85.0	86.4	84.4
IMDb movie reviews (2)	93.4	93.4	93.4	93.4	93.4	93.4	93.3	90.3	90.0	90.0	89.7	88.5	87.6
IMDb movie reviews (1)	90.7	90.7	90.7	90.7	90.6	90.7	90.5	89.4	98.4	89.4	89.2	85.1	87.3
Tweets SemEval	87.1	86.9	87.0	86.9	86.9	87.0	85.3	75.9	76.2	76.4	73.0	76.1	72.4
Tweets Airline	91.0	90.4	91.1	90.7	91.0	90.7	89.9	83.7	78.7	83.5	75.8	77.2	76.2
Sentiment140 (10%)	89.2	89.5	89.1	89.0	89.2	89.2	87.8	82.7	76.6	82.1	70.9	79.6	76.5

TABLE 7: COMPARING PRECISION FOR VARIOUS DATASET TYPES

Datasets	LC-S	CL-S	L-C	C-L	LTSM	CNN	SVM	LC-S	CL-S	L-C	C-L	LTSM	CNN
Music reviews	83.9	83.5	84.1	84.0	84.1	84.0	82.4	79.9	80.6	79.9	80.1	80.3	79.7

Book reviews	92.7	92.9	92.8	92.8	92.8	92.9	92.0	87.8	88.6	87.5
88.0	87.1	86.8								
Cornell movie reviews	91.8	91.7	91.8	91.9	91.8	91.9	91.2	85.7	85.6	86.2
85.0	86.4	84.4								
IMDb movie reviews (2)	93.4	93.4	93.4	93.4	93.4	93.4	93.3	90.3	90.0	90.0
89.7	88.5	87.6								
IMDb movie reviews (1)	90.7	90.7	90.7	90.7	90.6	90.7	90.5	89.4	98.4	89.4
89.2	85.1	87.3								
Tweets SemEval	87.1	86.9	87.0	86.9	86.9	87.0	85.3	75.9	76.2	76.4
73.0	76.1	72.4								
Tweets Airline	91.0	90.4	91.1	90.7	91.0	90.7	89.9	83.7	78.7	83.5
75.8	77.2	76.2								
Sentiment140 (10%)	89.2	89.5	89.1	89.0	89.2	89.2	87.8	82.7	76.6	82.1
70.9	79.6	76.								

By emphasizing the hybrid model's results, we can see that these models consistently produce the best results for every dataset. Word2vec or BERT solo models did not perform as well as hybrid models, which did. When Word2vec is used, the reliability results from hybrid models are better than those from single models. While these algorithms have attained a pretty good degree of precision, typically greater than 90%, the outcomes have also been enhanced using BERT, but less significantly.

Table .8.The AUC ratio for various dataset

Datasets	LC-S	CL-S	L-C	C-L	LTSM	CNN	SVM	LC-S	CL-S	L-C	C-L
LTSM	CNN										
Music reviews	83.9	83.5	84.1	84.0	84.1	84.0	82.4	79.9	80.6	79.9	
80.1	80.3	79.7									
Book reviews	92.7	92.9	92.8	92.8	92.8	92.9	92.0	87.8	88.6	87.5	
88.0	87.1	86.8									
Cornell movie reviews	91.8	91.7	91.8	91.9	91.8	91.9	91.2	85.7	85.6	86.2	
85.0	86.4	84.4									
IMDb movie reviews (2)	93.4	93.4	93.4	93.4	93.4	93.4	93.3	90.3	90.0	90.0	
89.7	88.5	87.6									
IMDb movie reviews (1)	90.7	90.7	90.7	90.7	90.6	90.7	90.5	89.4	98.4	89.4	
89.2	85.1	87.3									
Tweets SemEval	87.1	86.9	87.0	86.9	86.9	87.0	85.3	75.9	76.2	76.4	
73.0	76.1	72.4									
Tweets Airline	91.0	90.4	91.1	90.7	91.0	90.7	89.9	83.7	78.7	83.5	
75.8	77.2	76.2									
Sentiment140 (10%)	89.2	89.5	89.1	89.0	89.2	89.2	87.8	82.7	76.6	82.1	
70.9	79.6	76.5									

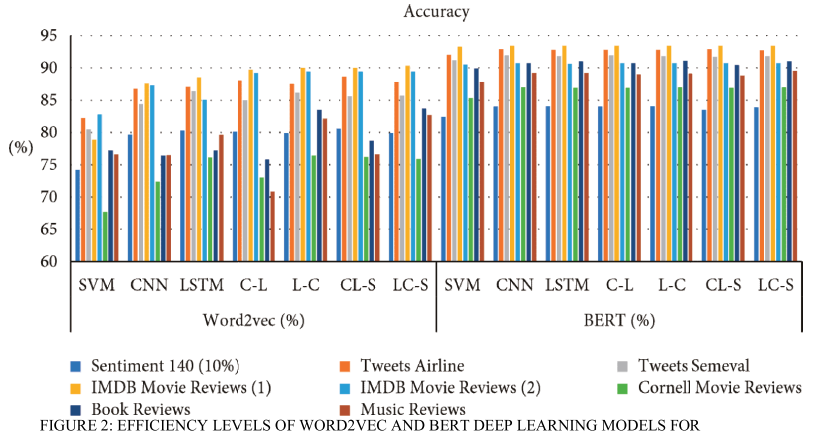

FIGURE 2: EFFICIENCY LEVELS OF WORD2VEC AND BERT DEEP LEARNING MODELS FOR VARIOUS DATASETS

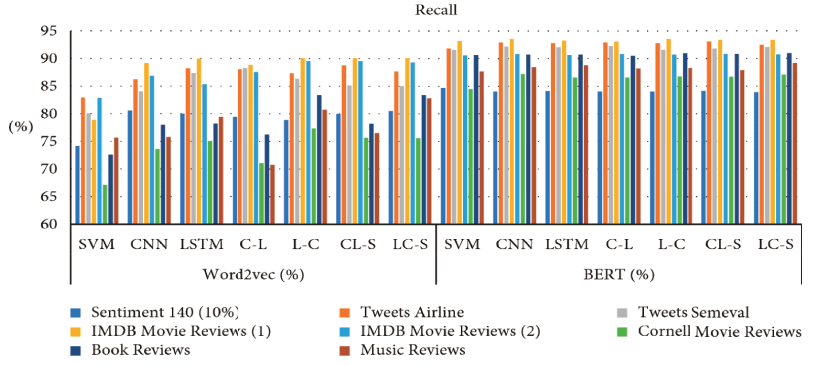

FIGURE 3: DEEP LEARNING MODEL RECALL VALUES USING WORD2VEC AND BERT FOR VARIOUS DATASETS

Given that reviews typically contain longer texts than tweets do, LCNN-SVN outperforms other hybrid models on larger linguistic samples (Table 4). While looking at the distribution of sample text lengths in some databases, the size of reviews ranges from 1 to 800 words. The Cornell movie reviews, meanwhile, are only 1 to 50 words long. A tweet can be between one and forty words long, yet the variation in sample length in the Sentiment140 dataset is right-skewed. It has been noticed that two databases, Sentiment140 and Cornell movie reviews, have poorer findings than the other data sources.

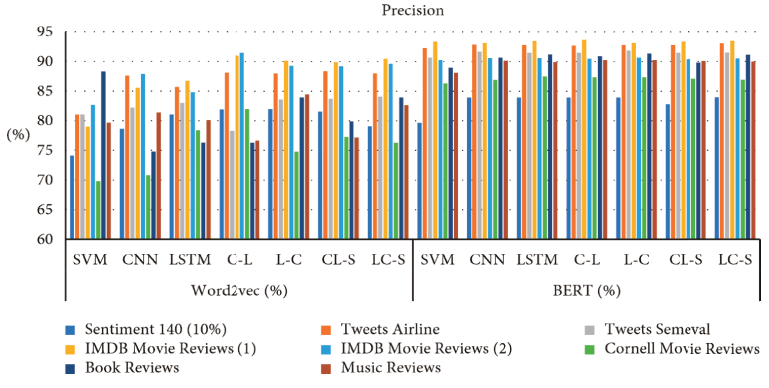

FIGURE 4: ACCURACY LEVELS OF WORD2VEC AND BERT DEEP LEARNING MODELS FOR VARIOUS DATASETS

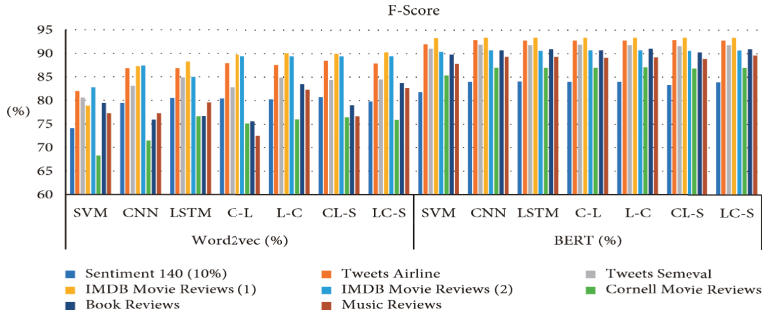

FIGURE 5: F-SCORE VALUES FOR DEEP LEARNING MODELS USING WORD2VEC AND BERT FOR VARIOUS DATASETS.

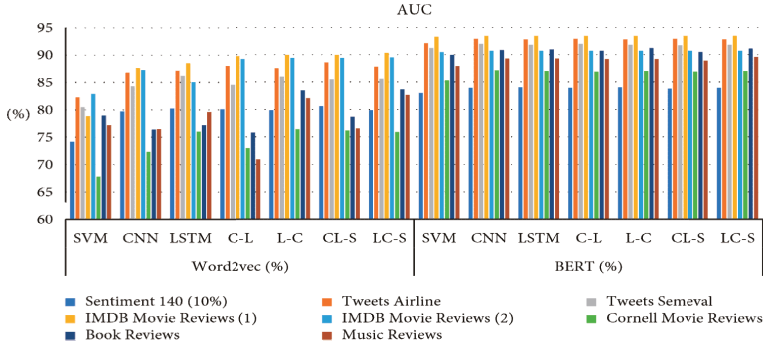

FIGURE 6: DEEP LEARNING MODEL AUC VALUES USING WORD2VEC AND BERT FOR VARIOUS DATASETS

In this reference, we offer further studies using a single dataset of tweets or reviews to classify sentiment. Remember that such hybrid models provide significantly better operation speed and precision outcomes. Eight different datasets were also used to assess the total effectiveness of such hybrid models, providing an unbiased evaluation of their total efficiency [18].

6. CONCLUSIONS

In this research, we suggested using social media network hybrid deep learning models data in Sentiment analysis. Using the Word2vec and BERT two-word extraction methods, we looked at how well SVM, CNN, and LSTM work together on eight literary comment and rating databases. Following that, we compared various hybrid models to single models. These studies aim to investigate the adaptability of hybrid models, specifically if hybrid methods can adjust to multiple database kinds and sizes. We investigated the effects of different dataset types, techniques for extracting features, and deep learning models on the accuracy of sentiment polarity analysis. Our tests show that, for sentiment polarity analysis, hybrid models consistently outperform all other models. Integrating deep learning models with the SVM method yields more accurate outcomes than doing it alone for conducting sentiment analysis. Hybrid models utilizing SVM are more reliable than those without it in most of the databases evaluated; nevertheless, their computing time is substantially greater. Researchers say that the datasets' quality and quantity greatly affected how well the techniques worked. Researchers are worried that the context of the information can greatly affect the choice of sentiment analysis models to get a better understanding of a particular subject, like trade, advertising, or healthcare, we plan to investigate the effectiveness of hybrid methodologies for sentiment analysis on hybrid databases and numerous hybrid contexts. Its relevance arises from linking sentiments to pertinent circumstances to offer customers individually tailored suggestions and feedback in great detail.

REFERENCES

1.N. C. Dang, M. N. Moreno-Garc´ıa, and F. De la Prieta, “Sentiment analysis based on deep learning: a comparative study,” Electronics, vol. 9, no. 3, p. 483, 2020.

2. A. Hassan and A. Mahmood, “Deep learning approach for sentiment analysis of short texts,” in :ird International Conference on Control, Automation and Robotics (ICCAR), IEEE, Nagoya, Japan, April 2017.

3. J. Qian, Z. Niu, and C. Shi, “Sentiment analysis model on weather related tweets with deep neural network,” in Proceedings of the 2018 10th International Conference on Machine Learning and Computing, ACM, Zhuhai, China, February 2018.

4. R. Alfrjani, T. Osman, and G. Cosma, “A hybrid semantic knowledgebase-machine learning approach for opinion mining,” Data & Knowledge Engineering, vol. 121, pp. 88–108, 2019.

5. I. Gupta and N. Joshi, “Enhanced twitter sentiment analysis using hybrid approach and by accounting local contextual semantic,” Journal of Intelligent Systems, vol. 29, no. 1, pp. 1611–1625, 2019.

6. H. Srinidhi, G. Siddesh, and K. Srinivasa, “A hybrid model using MaLSTM based on recurrent neural networks with support vector machines for sentiment analysis,” Engineering and Applied Science Research, vol. 47, no. 3, pp. 232–240, 2020.

7.Paulraj D, Ezhumalai P, Prakash Mohan (2024) A deep learning modifed neural network (DLMNN) based profcient sentiment analysis technique on twitter data. J Exp TheorArtifIntell 36(3):415–434.

8. Q.-H. Vo, H.-T. Nguyen, B. Le, and M.-L. Nguyen, “Multichannel LSTM-CNN model for Vietnamese sentiment analysis,” in 2017 9th international conference on knowledge and systems engineering (KSE), IEEE, Hue, Vietnam, October 2017.

9. A. U. Rehman, A. K. Malik, B. Raza, and W. Ali, “A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis,” Multimedia Tools and Applications, vol. 78, no. 18, pp. 26597–26613, 2019.

10. Z. Kastrati, L. Ahmedi, A. Kurti et al., “A deep learning sentiment analyser for social media comments in low-resource languages,” Electronics, vol. 10, no. 10, p. 1133, 2021.
11. N. Jnoub, F. Al Machot, and W. Klas, “A domain-independent classification model for sentiment analysis using neural models,” Applied Sciences, vol. 10, no. 18, p. 6221, 2020.
12. A. H. Ombabi, W. Ouarda, and A. M. Alimi, “Mining, “deep learning CNN–LSTM framework for Arabic sentiment analysis using textual information shared in social networks,” Social Network Analysis and Mining, vol. 10, no. 1, pp. 1–13, 2020.
13. J. F. Sanchez-Rada and C. A. Iglesias, "CRANK: a hybrid' model for user and content sentiment classification using social context and community detection," Applied Sciences, vol. 10, no. 5, p. 1662, 2020.
14. Samih Amina, GhadiAbderrahim, Fennan Abdelhadi (2022) Enhanced sentiment analysis based on improved word embeddings and XGboost. Int J ElectrComputEng 13:2.
15. Wang Q, Zhang W, Lei T, Cao Y, Peng D, Wang X (2023) CLSEP: contrastive learning of sentence embedding with prompt. KnowlBasedSyst 266:110381.
16. Ouni, C., Benmohamed, E., &Ltifi, H. (2024). Sentiment analysis deep learning model based on a novel hybrid embedding method. *Social Network Analysis and Mining*, *14*(1), 210.
17. Mutinda J, Mwangi W, Okeyo G (2023) Sentiment analysis of text reviews using lexicon-enhanced Bert embedding (LeBERT) model with convolutional neural network. Appl Sci 13:1445. https://doi.org/10.3390/app13031445.
18. “Sentiment140 - a twitter sentiment analysis tool,” Available from: (accessed on 10 December 2020), http://help. sentiment140.com/site-functionality.
19. K.-X. Han, W. Chien, C.-C. Chiu, and Y.-T. Cheng, “Application of support vector machine (SVM) in the sentiment analysis of twitter DataSet,” Applied Sciences, vol. 10, no. 3, p. 1125, 2020.
20. F. Abid, M. Alam, M. Yasir, and C. Li, “Sentiment analysis through recurrent variants latterly on convolutional neural network of Twitter,” Future Generation Computer Systems, vol. 95, pp. 292–308, 2019.
21. Siddiqui, S. T., Singha, A. K., Ahmad, M. O., Khamruddin, M., & Ahmad, R. (2022). IoT devices for detecting and machine learning for predicting COVID-19 outbreak. In Recent Trends in Communication and Intelligent Systems: Proceedings of ICRTCIS 2021 (pp. 107-114). Singapore: Springer Nature Singapore.
22. Singha, A. K., Zubair, S., Malibari, A., Pathak, N., Urooj, S., & Sharma, N. (2023). Design of ANN Based Non-Linear Network Using Interconnection of Parallel Processor. Computer Systems Science & Engineering, 46(3).
23. Zubair, S., Singha, A. K., Pathak, N., Sharma, N., Urooj, S., &Larguech, S. R. (2023). Performance Enhancement of Adaptive Neural Networks Based on Learning Rate. Computers, Materials & Continua, 74(1).
24. Singha, A. K., & Zubair, S. (2023). Combination of Optimization Methods in a Multistage Approach for a Deep Neural Network Model. International Journal of Information Technology, 1-7.
25. Singha, A. K., Pathak, N., Sharma, N., Tiwari, P. K., & Joel, J. P. C. (2022). COVID-19 Disease Classification Model Using Deep Dense Convolutional Neural Networks. In Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2022, Volume 2 (pp. 671-682). Singapore: Springer Nature Singapore.
26. Singha, A. K., Pathak, N., Sharma, N., Tiwari, P. K., & Joel, J. P. C. (2022). Forecasting COVID-19 Confirmed Cases in China Using an Optimization Method. In Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2022, Volume 2 (pp. 683-695). Singapore: Springer Nature Singapore.

[1]ANJANI KUMAR SINGHA
DEPARTMENT OF SCHOOL OF COMPUTER SCIENCE AND ENGINEERING, GALGOTIA UNIVERSITY, GREATER NOIDA
Email address- anjani348@gmail.com

[2]SHAMS SIDDQUI
DEPARTMENT OF COMPUTER SCIENCE, JAZAN UNIVERSITY, SAUDI ARABIA
Email address - stabrez@jazanu.edu.sa

[1]PRADEEP KUMAR TIWARI
DEPARTMENT OF COMPUTER APPLICATIONS, DR. VISHWANATH KARAD MIT WORLD PEACE UNIVERSITY, PUNE, MAHARASHTRA, INDIA
DELHI-NCR CAMPUS, GHAZIABAD, 201204, INDIA
Email address- pradeeptiwari.mca@gmail.com

[4]*GAURAV YADAV
DEPARTMENT OF COMPUTER SCIENCE, ALIGARH MUSLIM UNIVERSITY, ALIGARH UTTAR PRADESH, INDIA
Email address - gyadav444@gmail.com
(Corresponding Author)

[5]SURYA KANT
DEPARTMENT OF MASTER OF COMPUTER APPLICATION, SHRI RAM GROUP OF INSTITUTIONS, JABALPUR, MP
Email address- suryak@srmist.edu.in