Journal of Discrete Mathematical Sciences & Cryptography
ISSN 0972-0529 (Print), ISSN 2169-0065 (Online)
Vol. 29 (2026), No. 1, pp. 375–396
DOI : 10.47974/JDMSC-2633

Sentiment analysis on social-media textual data using hybrid transformer and sequence models

Harihar Nath Verma [§]
Sanjay Kumar Malik [†]
University School of Information, Communication and Technology
Guru Gobind Singh Indraprastha University
Dwarka 110078
Delhi
India

Vanita Jain *
Department of Electronics and Communication Engineering
Faculty of Technology
University of Delhi
Delhi 11007
India

Abstract

Sentiment analysis has become an invaluable exercise on social media addressing the understanding of opinion and consumer behavior and trends. The system of informality and context richness against platforms, such as Twitter, raises serious issues to current sentiment classification models. It is hypothesized that integration of the contextual feature representation capacity of embeddings in transformers with the sequential learning ability of BiLSTM networks in DitilBERT-BiLSTM can help to improve accuracy in sentiment classification. The Sentiment140 dataset is used as a training and testing set with 1.6 million small tweets among which there are equal numbers of positive and negative tweets. Experiments demonstrate that the proposed model obtains superior performance compared to traditional deep learning architectures, 81% accuracy, 82% precision, recall of 80% and F1-score of 81%. DistilBERT-BiLSTM shows good generalization and learns efficiently using less computer power than full-scale transformer models, indicating that it is more suitable to applications of real-world social media sentiment analysis. This study points out the success of hybrid deep learning networks in modelling both global semantic structure and local

[§] *E-mail:* **hnvermagwl@gmail.com**
[†] *E-mail:* **skmalik@ipu.ac.in**
* *E-mail:* **vanitajain@fot.du.ac.in (Corresponding Author)**

sequential relations as a strong remedy to resolving opinion mining in dynamic and noisy digital situation.

Subject Classification: *68T07, 68T50.*

Keywords: *Sentiments, Tweets, Social media text, DistilBERT, BiLSTM, Hybrid deep learning model, Transformer-based embeddings, Natural language processing, Sentiment140 dataset, Opinion mining.*

1. Introduction

Over the recent years, Twitter, Instagram, Facebook and many blogging websites have experienced an explosive rise in user activity. These online environments can enable people to confidently say out their thoughts, feelings and perspectives on a broad variety of topics, which makes the constant creation of huge quantities of unstructured textual information a reality [1]. With over 500 million active users in India alone, social media serve as a massive, real time source of user-generated textual data. Extracting actionable insights from this dynamic and unstructured content presents a significant challenge and opportunity, especially in domains such as political analysis, brand reputation management, public health monitoring and crisis response. Social media platforms, like Twitter, are now widely used by everyone, where everyone posts their opinions regarding products, ongoing trends, politics etc. Texts' emotional tones may be determined via sentiment analysis. Twitter's sentiment analysis may be quite useful in determining public opinion on current events, trends and other subjects [2,3,4]. Sentiment analysis includes three main steps pre-processing, feature extraction and classification. During pre-processing, the text is cleaned, i.e. the stop-words, special characters and numbers are removed. Techniques, such as, Term Frequency – Inverse Document Frequency (TF-IDF), Global Vectors for Word Representaion (GloVe), fastText or word2vec are then used to generate features. Last, machine learning is used to categorize the sentiment [5]. The informal language of social media, such as slang, abbreviations, emojis, sarcasm or even short sentences also makes it more challenging to interpret sentiments correctly [6]. Enormous improvement in the efficacy and precision of sentiment analysis has occurred with the introduction of transformer-based language models like DistilBERT, Robustly Optimized BERT-pretraining Approach (RoBERTa) and Bidirectional Encoder Representation from Transformer (BERT) [7,8]. These models offer contextual word embeddings that understand the meaning of words based on their

surrounding context, enabling more accurate sentiment classification even in short and ambiguous texts. In this study, a hybrid deep learning (DL) model is put forward that combines the contextual embedding capability of DistilBERT and the sequential learning capacity of Bidirectional Long Short-Term Memory (BiLSTM) in an attempt to enhance the classification of sentiments on contextual content in social media. DistilBERT-BiLSTM model is tested on Sentiment140 dataset [9], which is the benchmark corpus of 1.6 million tweets labelled with positive or negative. DistilBERT is selected since it is computationally efficient and performs comparatively better with BERT, hence fit in real-time and constrained resource applications [10]. Whereas there have been previous efforts to examine transformer-based or sequential models individually, the hybrid system proposed here seeks to unite their best qualities [11]. This aims at coming up with a light but precise sentiment analysis system that can process the complexity of social media text at a large scale. The proposed model improves traditional classifiers in sentiment search activities, according to F1-score, recall, accuracy and precision.

A. Aim and contribution

An effective and efficient sentiment analysis system for textual data from social media platforms is the main focus of this paper. DistilBERT-BiLSTM, a hybrid model, is used for this purpose. This model combines the lightweight transformer architecture of DistilBERT with the sequential processing power of BiLSTM, enabling better contextual understanding and handling of sequential dependencies in short-form, informal texts, such as tweets.

The major contributions of this paper are outlined below:

- Utilized the Sentiment140 dataset [9] comprising 1.6 million real-world tweets, ensuring data diversity and large-scale evalution.
- Implemented DistilBERT-BiLSTM, a hybrid model integrating transformer embeddings with BiLSTM sequence learning.
- Applied extensive text pre-processing techniques including tokenization, lemmatization, stop-word removal and normalization.
- Achieved state-of-the-art performance in binary sentiment classification with 81% accuracy, outperforming traditional and some deep learning methods.

- Conducted detailed performance evaluation using accuracy, precision, recall, F1-score, confusion Metric and loss/accuracy curves.

B. *Structure of paper*

The following is the structure of the paper: The work on sentiment analysis in textual data from social media is reviewed in Section II. Data collection, pre-processing and the creation of the DistilBERT-BiLSTM model are all covered in depth in Section III. Results and comparisons are shown in Section IV. Important takeaways and suggestions for further study are included in Section V.

II. Literature review

A thorough literature review is conducted to obtain a proper understanding of applied machine learning (ML) models in the area of social media textual data. Prior research has explored various techniques, revealing their strengths and limitations, that summarized in Table 1.

Wang, Wei, and Tian [12] conducted a comprehensive comparative analysis of ten ML models for sentiment classification of news media articles related to transboundary rivers. Using a dataset of 9,382 articles spanning from 1977 to 2022, the study evaluated models such as the BERT model, demonstrated the highest overall performance achieving an accuracy of 72%, outperforming traditional models and the Adaptive Fuzzy Inference Neural Network (AFINN) sentiment dictionary. BERT is particularly effective in identifying conflict-related sentiments and showed exceptional predictive.

Golubev, Rusnachenko, and Loukachevitch [13] presented RuSentNE-2023, a benchmark evaluation focused on targeted sentiment analysis in Russian news texts. The task involves identifying sentiments towards named entities within single sentences using the RuSentNE corpus, which includes annotations for entities, sentiment polarity, emotional states and effects. The top-performing model achieved a macro F1-score of 66.67%, while Chat Generative Pre-trained Transformer (ChatGPT), used in a zero-shot setting, achieved an impressive 60%, ranking 4th among participants.These findings indicate the interpolating nature of sentiment analysis in establish news contexts and the prospects of transformer-based architectures and ensemble approaches in regards to enhancing performance.

Tang [14] investigates the performance of transformer neural networks in sentiment analysis over the social media data on a large scale. Based on the Sentiment140 dataset [9] containing 1.6 million tweets, the paper contrasts transformer with a standard and deep-learning method, such as Support Vector Machine (SVM), Lont Short Term Memory (LSTM) (2 and 3 layers), BiLSTM and types of Recurrent Neural Network (RNN). The findings indicate that the Transformer model is the best model in classification accuracy and robustness, outperforming all the baselines, which is quite remarkable in pre-trained model capabilities to process sequential data and detect sentiment patterns in user-generated data. The proposed transformer-based model outperforms the best competing baseline, LSTMx3, which has an accuracy of 76.5% of the proposed transformer-based model at classifying online sentiment tasks.

Sharma and Bhalla [15] described a dual contrastive learning (DCL) solution to hate speech analysis of code-mixed social media text Hindi-English, a linguistically demanding setting more common in multilingual countries such as India. By integrating multilingual sentBERT representations with sentence-based records of the Universal Sentence Encoder, the suggested model is very efficient in representing the sentiment polarity. TF-IDF is used to enhance robustness and performance of the classification using the DCL. Experiments conducted using traditonal ML models like KNN, DT and logistic regression with results showing 73.63%, 79.78%, and 85.50% accuracy respectively. On average, the DCL model outperforms baseline methods by achieving 86% accuracy.

Ilham et al. [16] explored the impact of sarcasm detection on the performance of sentiment analysis in Indonesian-language tweets. Using a dataset of 1,000 tweets, the study applies Random Forest for sarcasm detection and NB with TF-IDF features for sentiment classification. The results reveal that incorporating sarcasm detection does not enhance sentiment classification accuracy. The baseline sentiment analysis achieved an accuracy of 74.3%, while sentiment analysis with sarcasm detection slightly declined to 72.9%. The sarcasm detection model itself recorded an average accuracy of 71.7%, indicating moderate performance in identifying sarcastic content.

Handoyo et al. [18] proposed a contextual sarcasm detection model on Twitter data using RoBERTa with word embedding-based data augmentation. On the highly imbalanced iSarcasm dataset [17], data augmentation resulted in an F1-score improvement from 37.2% to 40.4% (+3.2%), accompanied by a corresponding increase in the Matthews Correlation Coefficient (MCC) to 0.3084. However, on larger and balanced

datasets like Ghosh and Ptacek, performance gains are minimal or negative, highlighting the sensitivity of augmentation techniques to dataset structure and balance [18].

Table 1

Overview of sentiment analysis of social media text using machine and deep learning models

Author	Methodology	Data	Key Findings	Limitation/Future Work
Wang, Wei, and Tian [12]	10 ML models (KNN, NB, SVM, DT, RF, GBDT, XGBoost, MLP, LSTM, BERT)	9,382 news articles on transboundary rivers (1977–2022)	BERT (72.7%) outperformed all; effective in conflict sentiment in the Indus Basin	Focused on 5-class sentiment only; expansion to multilingual datasets suggested
Golubev, Rusnachenko, and Loukachevitch [13]	Transformer-based, RuBERT, ChatGPT (zero-shot)	RuSentNE-2023: Russian News Corpus	Best model: F1-macro = 66.67%; ChatGPT (zero-shot) = 60% (rank 4th)	Language and news-specific; generalizability to other languages not explored
Tang [14]	Transformer, compared with SVM, LSTMx2, LSTMx3, BiLSTM, RNN	Sentiment140 (1.6M tweets)	Transformer outperformed all, beating LSTMx3 (76.5%) and others in robustness and accuracy.	Needs evaluation across multilingual or low-resource languages
Sharma and Bhalla [15]	ML: sentBERT + TF-IDF	Code-mixed Hindi-English social media data	Accuracy: KNN (73.63%), DT (79.78%), LR (85.50%); DCL outperformed by achieving 86%	Code-mixing complexity may vary across datasets.
Ilham et al. [16]	Sarcasm: Random Forest; Sentiment: Naïve Bayes with TF-IDF	1,000 Indonesian tweets	Sentiment accuracy dropped from 74.3% to 72.9% with sarcasm detection; sarcasm model: 71.7%	Sarcasm detection didn't enhance sentiment performance.

Contd...

Handoyo et al. [18]	RoBERTa + GloVe-based data augmentation for sarcasm detection	Sarcasm, Ghosh, Ptacek, SemEval-18	Sarcasm F-score ↑ from 37.2% to 40.4% (+3.2%); gains are minimal in large balanced.	Augmentation is useful for imbalanced data only; mixed results elsewhere.

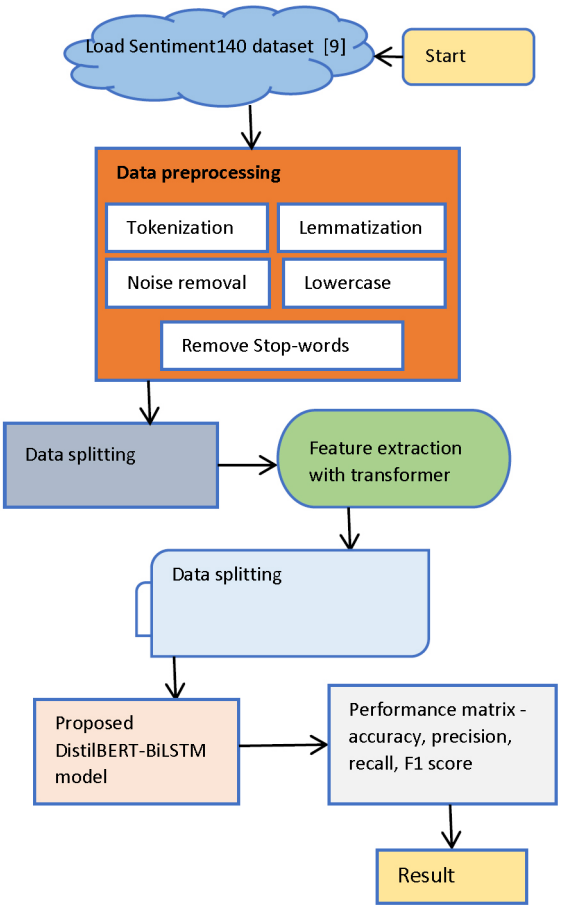

Figure 1

Flowchart for Social Media textual data analysis

Figure 1 shows the whole structure of the sentiment analysis on textual data from social media, with an emphasis on the steps taken from obtaining the raw data to reaching a final judgement.

III. Methodology

This study follows a structured and well-defined pipeline for sentiment analysis, as shown in Figure 1. The process begins with data acquisition using the Sentiment140 dataset [9], which contains 1.6 million tweets equally divided into positive and negative sentiments. Its large-scale and real-world composition provides a robust benchmark for evaluating sentiment models in dynamic social media environments. In the pre-processing stage, raw tweets are cleaned and standardized to handle the informal nature of social media language. The most significant steps involve tokenization, lemmatization, lowercasing, removal of stop-words and special characters. DistilBERT tokenizer is employed in text encoding with semantics maintained, and this can be applied to the transformer-based architecture. The data is then cleaned and tokenized and sent into the DistilBERT transformer to produce a contextual embedding that represents deep semantic content in the input text. This is followed by these embeddings being run through a BiLSTM that has the capability of reading sequences forward and backward to know neighbour dependencies and sequential orientation. The architecture of this hybrid model DistilBERT-BiLSTM can grasp both global context and word order patterns of the sentences and is, therefore, an excellent choice to perform sentiment analysis on short text.

A. *Data Collection*

The dataset that is used in this research activity is Sentiment140 dataset [9], which is a common benchmark in sentiment analytics in social media. It includes 1.6 million of tweets brought together through the Twitter API with annotations encompassing distant supervision using emoticons within the tweet text.

The tweets are labelled with binary sentiment classes:

0 for negative sentiment

1 for positive sentiment

In classifying the responses as binary, 1 is remapped to 1, translating the data set to a clean 0-1 binary representation more convenient to deep learning algorithms. The final dataset used in this work is perfectly balanced, with 800,000 positive tweets and 800,000 negative tweets, ensuring unbiased training and evaluation across sentiment classes.

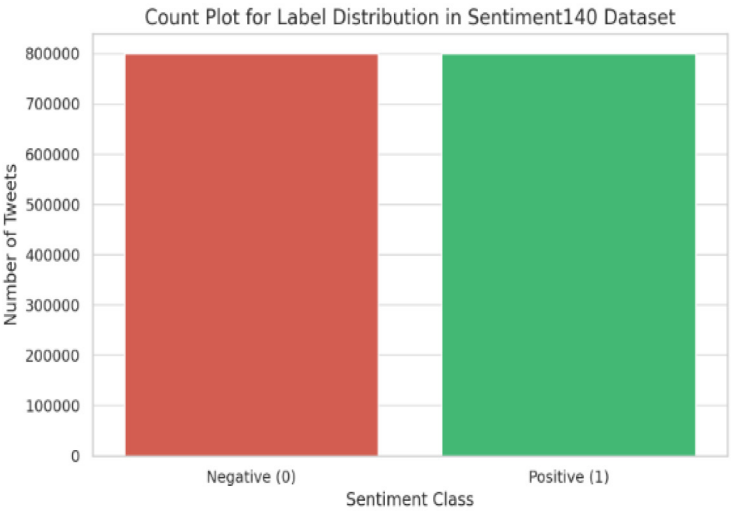

Figure 2

Count plot for label distribution of data.

Figure 2 presents the count distribution of the sentiment classes in the dataset. The balanced nature of the dataset is visually evident, with equal frequencies for both positive, shown in green, and negative, shown in red, sentiment labels. This balanced structure is particularly valuable for training classification models without the risk of class bias, leading to more stable and generalizable performance.

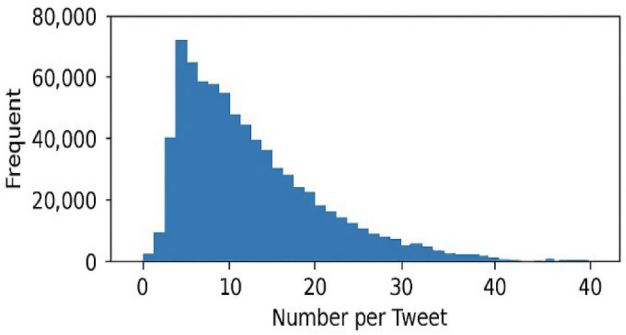

Figure 3

Sample Tweet Length Distribution

Figure 3 displays the distribution of tweet lengths in the dataset. Most tweets are short, with an average length of approximately 15 words,

consistent with Twitter's 280-character limit. The brevity and informal language of these tweets present a unique challenge for sentiment classification, making the use of deep contextual embeddings like DistilBERT especially relevant.

The dataset also includes meta-fields such as: (Described in Table 2)

- Tweet ID
- Timestamp
- User handle
- Query flag
- Cleaned tweet text

TABLE 2

Sentiment140 Dataset fields with description

Field	Description
Tweet ID	Unique identifier for each tweet
Timestamp	Date and time when the tweet was posted
User-handle	Username of the tweet's author
Query flag	Indicate whether a search query provided
Cleaned tweet text	Text of tweet after pre-processing
Sentiment label	Associated sentiment class (0 or 1)

This rich structure provides not only sentiment labels but also temporal and user-based insights, making it ideal for sentiment analysis research in dynamic online environments.

B. *Data Preprocessing*

A text pre-processing technique like to the one outlined in the text must be used after the data collection phase in order for a model to correctly classify textual input [19,20]. The text data has to be pre-processed using techniques including tokenization, lemmatization, noise removal and case normalization, before being input into machine learning models. At this stage, a number of critical procedures are executed to clean up and standardize the text data, such as:

- **Tokenization:** At this stage, the original text is divided into tokens. Tokenisation is the process of breaking down the text into its component words.
- **Lemmatization:** The process of reducing a word to its most basic form involves removing or replacing the suffix at this stage. It also entails reducing the count of distinct occurrences of related phrases.
- **Noise Removal:** Among other things, spaces, punctuation, numbers, email addresses and links are deleted to make room for the desired voice.
- **Lowercase:** Drop the capitalisation and change every word to lowercase. What makes ML model training more difficult is because ML models consider "Real" and "real" to be independent terms. Which is why, to get around this problem, all the letters are converted into lowercase.
- **Stop Word Removal:** The purpose of stop words like she, them, their, us, etc. is to aid human understanding but they do not contribute to the content of the phrase. The learning feature set is therefore made easier by removing these stop words.

C. *Data Splitting*

Data splitting involves dividing the dataset into separate subsets to train and evaluate a machine learning model. In this, an 80:20 split is made, means 80% of the data is used for training the model, while 20% is reserved for testing its performance.

D. *Proposed DistilBERT-BiLSTM model*

The DistilBERT-BiLSTM model is a hybrid DL architecture that merges the strengths of DistilBERT's contextual word embeddings with the sequential processing power of BiLSTM networks, making it well-suited for text classification tasks. DistilBERT is a smaller, faster version of BERT that retains most of its language understanding capabilities while requiring fewer computational resources. On the other hand, BiLSTM networks can analyze data sequences from both past and future contexts, which is essential for capturing the meaning and flow in natural language.

The model takes as input a sequence of tokens represented as:

$$X = (x_1, x_2, \ldots, x_T) \tag{1}$$

Where T is the length of the sequence.

Step 1: Contextual Embedding via DistilBERT

The DistilBERT embedding layer transforms the input sequence X into a sequence of dense contextual embeddings:

$$H = (h_1, h_2, \ldots, h_T), \quad h_t \in R^d \tag{2}$$

Here, each embedding vector h_t encodes the semantic meaning of the token x_t considering the full context of the sentence. The dimension of each embedding is d.

Step 2: Bidirectional LSTM Layer

The BiLSTM consists of two separate LSTM networks:

- Forward LSTM processes the sequence from x_1 to x_t

$$h_t = \text{LSTMforward}(x_t, h_{t-1}) \tag{3}$$

- Backward LSTM processes the sequence from x_t to x_1:

$$h_t = \text{LSTMbackward}(x_t, h_{t+1}) \tag{4}$$

The hidden states at time t from both directions are concatenated to form the final BiLSTM output:

Step 3: Pooling Layer

To convert the sequence of BiLSTM outputs into a fixed-length vector h^{pool}, A pooling operation, such as max-pooling or attention pooling, is applied:

$$h_{pool} = \text{Pooling}(h_1^{bi}, h_2^{bi}, \ldots, h_T^{bi}) \tag{5}$$

Step 4: Fully Connected and Output Layer

The pooled vector h^{pool} is passed through one or more fully connected layers f(·):

$$z = f(h_{pool}) \tag{6}$$

Finally, a SoftMax activation function is applied to obtain class probabilities:

$$\hat{y} = SoftMax(z) \tag{7}$$

where $\hat{y}$ is the predicted probability distribution over the classes.

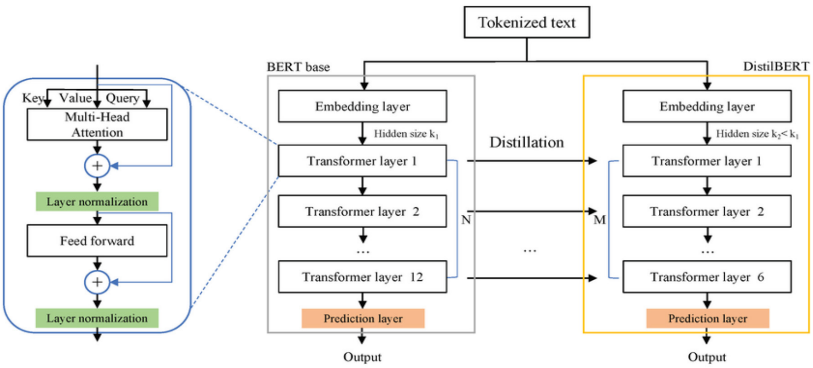

Figure 4

DistilBERT model Model Summary

Figure 4 illustrates the internal architecture of the proposed DistilBERT model, which is derived from the BERT-base model through knowledge distillation. The BERT-base model consists of 12 transformer layers with multi-head attention, layer normalization and feed-forward components, producing deep contextual embeddings for input text. DistilBERT, as a lighter and faster variant, retains the same embedding mechanism but reduces the number of transformer layers to six, effectively lowering computational complexity while preserving performance. The distillation process enables the DistilBERT model to learn from the BERT-base model by replicating its outputs using a distillation loss function. This architecture offers a significant speedup in training and inference time, making it highly suitable for large-scale and real-time sentiment analysis tasks such as those performed on the Sentiment140 twitter dataset [9].

E. *Performance metrics*

The models' performance is assessed using a range of metrics, such as accuracy, precision, etc. An additional statistic that delineates the efficacy of models is the confusion Metric. The usual way that ML and DL models are evaluated is by using a confusion Metric, which gives the results in terms of TP, TN, FP, and FN. For assessments, the following measures are used.

Accuracy: The percentage of test samples that the model properly classifies using a certain dataset. In equation (8).

$$Accuracy = \frac{TN + TP}{TP + TN + FP + FN} \tag{8}$$

Recall: The true-positive rate (TPR) is the percentage of positive samples that are really anticipated to be positive, relative to all true positive and false negative samples. In equation (9).

$$\text{Recall} = \frac{TP}{TP + FN} \tag{9}$$

Precision: The fraction of samples that actually tested positive out of all the samples that are anticipated to test positive. In equation (10).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{10}$$

F1 Score: An amalgamation of recall and precision. Both recall and precision are inversely correlated; as one rises, the other falls. By including the F1-score into equation (11) will bring these two measures into harmony.

$$F1 = \frac{2 * (\text{precision} * \text{recall})}{\text{precision} + \text{recall}} \tag{11}$$

These parameters are employed to evaluate model performance and compare with existing models.

IV. Results and discussion

This part evaluates the effectiveness of the suggested hybrid transformer-sequence frameworks in sentiment analysis on twitter-based textual data. The trial is conducted on a cloud environment (Google Colab Pro+), and the deep learning framework PyTorch is used. The use of GPU speeded up the training (with a Tesla T4) along with 16GB of RAM and an Intel Xeon processor on a virtual memory by supporting the usage of Linux whatever be the size of the data used to train the model. The Sentiment140 data [9] used in this analysis contains about 1.6 million tweets. The dataset used in this tweet dataset is balanced because more than 6000 tweets have been evenly divided into the two categories of a sentiment: positive and negative. Table 3 shows the outcomes of the different hybrid model, DistilBERT-BiLSTM. The four key measurements that are used to determine the models performance are Accuracy, Precision, Recall, and F1-score. DistilBERT-BiLSTM performed better than the others with 82% precision, an 81% accuracy, 80% recall, and 81% F1-score. The

model can capture the context and sequential patterns of tweet-based text data, and therefore an ideal candidate for sentiment analysis tasks on social media.

Table 3

Performance of Hybrid Models on Sentiment140 Dataset

Performance Metric	DistilBERT-BiLSTM
Accuracy	81.00
Precision	82.00
Recall	80.00
F1-score	81.00

Figure 5

Accuracy Curve of DistilBERT-BiLSTM Model

Figure 5 indicates the accuracy of training and validation of DistilBERT-BiLSTM model after 4 iterations of the epoch training. The model had repeated accuracy, and training and validation accuracy curves reached the optimal of 81% signifying high generalization and learning using the twitter sentiment data. The model showed rapid convergence within the first two epochs and maintained stability thereafter.

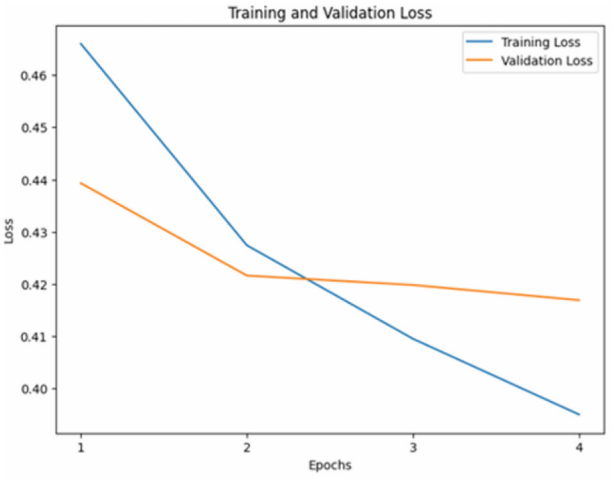

Figure 6

Loss Curve of DistilBERT-BiLSTM Model

Figure 6 displays the training and validation loss trends of the DistilBERT-BiLSTM model. Over the course of four epochs, a reduction in training loss from 0.47 to 0.40 and validation loss from 0.44 to 0.42 is seen; this pointed to effective optimisation. The close alignment between training and validation losses suggests excellent generalization on unseen data.

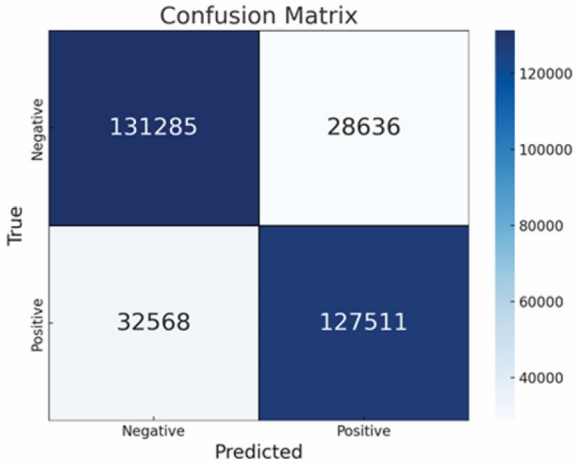

Figure 7

Confusion Metric for DistilBERT-BiLSTM Model

Figure 7 presents the confusion Metric for the DistilBERT-BiLSTM model, which correctly classified 131,285 negative and 127,511 positive instances. The model had only 28,636 false positives and 32,568 false negatives, highlighting its high accuracy and low error rate in binary sentiment classification. The Metric demonstrates that the model performs well in both precision and recall for both sentiment classes, validating its suitability for real-world social media analysis tasks.

A. *Comparison with traditional models*

A comparative performance study is carried out against standard baseline models, such as NB and SVM, in order to assess the efficacy of the suggested model. As indicated in Table 4, the comparison is based on four important performance metrics: recall, accuracy, precision, and F1-score. DistilBERT-BiLSTM hybrid model displays the best results out of the three methods with the accuracy of 81% as opposed to 75% on SVM and 72% on NB. This implies a dramatic rise in the model capacity to detect the sentiments in social media text accurately. Regarding precision, the compared model achieves 82%, higher than SVM (70%) and NB (71%), which is also strong in eliminating false positives. The SVM model achieves a slightly higher recall than the DistilBERT-BiLSTM (91 vs. 80) but it also provides better F1-score (81) balancing recall and precision better than the other methods (SVM: 79%, NB: 69). These findings are indicative of the greater ability of the hybrid model to explain complex and context-based informal textual data that is often encountered in mediums such as Twitter.

Table 4

Comparison between the other model and the proposed model

Metric	NB[21]	SVM[22]	BERT[23]	DistilBERT-BiLSTM
Accuracy	72	75	57	**81**
Precision	71	70	57	**82**
Recall	68	91	60	**80**
F1-score	69	79	57	**81**

The accuracy and F1-score of the suggested DistilBERT-BiLSTM model are highly competitive with those of the traditional methods, and in addition, the model demonstrates desirable equilibrium between timing-overall contextual comprehension provided by DistilBERT and sequential stream of sentiments delivered by BiLSTM. This can be very

powerful when using real-time sentiment analysis over a noisy unstructured social media text, with obvious benefits during actual deployments.

B. *Justification and Novelty*

DistilBERT-BiLSTM hybrid model is selected as it does a great job synthesizing two strong suits: the context learning capabilities of the transformer models and the capacity of BiLSTM networks to encode sequential regularities. The informal, slang-based language of social media, full of abbreviations and sarcasm, may complicate the more traditional and more straightforward models. Standalone transformers occasionally fail to capture the subtlety of sequential hints, yet using the lightweight and pre-trained transformer representations of DistilBERT augmented with sequential memory of BiLSTM can help retain word order and relations in text. The novelty of the work is the construction of a model, which is not only computationally efficient but also has high accuracy, trained and tested on large-scale, real-world twitter data Sentiment140 [9], which has 1.6 million tweets in it. DistilBERT-BiLSTM achieved accuracy results of 81% which is above the baseline model, which included standalone LSTM, Convolutional Neural Networks (CNNs) and classical classifiers. Also, its lower training time, high precision and recall present it as an inviting, scalable approach to real-time sentiment analysis in dynamic digital contexts.

C. *Limitations and Future Scope*

In spite of its positive outcomes, the model has several limitations. The lighter-weight DistilBERT-BiLSTM still requires GPU acceleration to train and execute effectively, which radiates its usability bearing on low-resource gadgets or mobile devices without pruning or quantization approaches. The other limitation is that the model is trained on English tweets alone, making its ability on multilingual or domain-specific sentiment evaluation an open question. Also, because of the balancedness of the dataset, the robustness of the model to real-world imbalanced data situations is not referred.

Potential future improvements include:

- Extending support to multiple languages by integrating multilingual transformers.

- Creating lightweight versions optimized for mobile and edge devices via pruning or knowledge distillation.
- Enabling real-time sentiment detection through live stream processing of social media trends.
- Improving transparency and trustworthiness with explainable AI (XAI) methods to better interpret model predictions.
- Expanding the scope from binary sentiment to multi-label or emotion-based sentiment classification.

V. Conclusion

Due to the influx of user-generated content in sites such as Twitter, sentiment analysis has become a vital tool that allows a snippet of how people feel, market trends, and behavior of users. This research proves that DistilBERT-BiLSTM hybrid model helps to properly identify the tone of Twitter text with high precision, which can be used as an alternative in sentiment analysis of short-text since most methods are not effective in relating the short-text. With the high accuracy of 81%, excellent precision and recall, this is possible due to the complementary design with DistilBERT managing holistic contextual knowledge and BiLSTM maintaining word-in-sequence features. This type of combination enables the model to understand minor linguistic details more effectively than that of single architectures. In prospect, the study opens the doors of real-time, multilingual, and customizable sentiment analysis solutions. With currently increasing volume and complexity of social media content, such hybrid models will be increasingly important to addressing the challenges of large-scale, delicate text understanding.

In the future, there are a few areas in which it can be further enhanced. These involve scaling the model to handle multilingual data with the help of large transformers, creating more efficient versions to work on mobile and edge hardware, supporting sentiment analysis on streaming data, and making models more transparent through explainable AI. Further, going beyond binary sentiment classification towards multi-label and emotion-based sentiment analysis would result in a more general system and more potential use cases.

References

[1] L. Chaudhary, N. Girdhar, D. Sharma, J. Andreu-perez, and A. Doucet, "A Review of Deep Learning Models for Twitter Sentiment Analysis: Challenges and Opportunities", *Journal of Latex Class Files*, vol. 14, no. 8, pp. 1–30 (2015).

[2] H. Pal and B. Bhushan, "Sentiment Analysis on Twitter Dataset using Voting Classifier", International Conference on Electrical Electronics and Computing Technologies (ICEECT), India, pp. 1–6 (Aug. 2024), doi: 10.1109/ICEECT61758.2024.10739316.

[3] M. R. Baker, and A. Utku, "Unraveling user perceptions and biases: A comparative study of ML and DL models for exploring twitter sentiments towards ChatGPT", *Journal of Engineering Research,* vol. 13, no. 2, pp 1658–1665 (2025), doi: 10.1016/j.jer.2023.11.023.

[4] A. Jain, and V. Jain, "Sentiment classification of twitter data belonging to renewable energy using machine learning", *Journal of Information and Optimization Sciences*, vol. 40, no. 2, pp. 521–533 (Mar. 2019). doi: 10.1080/02522667.2019.1582873.

[5] K. L. Tan, C. P. Lee, and K. M. Lim, "A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research", *Applied Sciences*, vol. 13, no. 7, pp 1–21 (2023), doi: 10.3390/app13074550.

[6] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis", *Foundations and Trends in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135 (2008), doi: 10.1561/1500000011.

[7] J. Devlin, M. Chang, K. Lee, and K. Tautanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", *Computation and Language* (May 2019). doi: 10.48550/arXiv.1810.04805.

[8] A. A. Khan, S. K. Malik, and V. Jain, "Semantic search framework over knowledge bases using embeddings-based similarity", *Journal of Discrete Mathematical Sciences & Cryptography,* vol. 27, No. 6, pp. 1963–1975 (Sep. 2024). doi: 10.47974/JDMSC-2047.

[9] https://www.kaggle.com/datasets/kazanova/sentiment140.

[10] Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter", *Computation and Language*, pp 1–5 (Mar. 2020). doi: 10.48550/arXiv.1910.01108.

[11] V. Jain, P. S. Lamba, B. Singh, N. Namboothiri, and S. Dhall, "Facial expression recognition using feature level fusion", *Journal of Discrete*

Mathematical Sciences and Cryptography, vol. 22, no. 2, pp. 337–350 (Mar. 2019). doi: : 10.1080/09720529.2019.1582866.

[12] J. Wang, J. Wei, and F. Tian, "A comparative study of machine learning models for sentiment analysis of transboundary rivers news media articles", *Soft Computing*, vol. 28, no. 23, pp. 13331–13347 (Dec. 2024). doi: 10.1007/s00500-024-10357-2.

[13] Golubev, N. Rusnachenko, and N. Loukachevitch, "RuSentNE-2023: Evaluating Entity-Oriented Sentiment Analysis on Russian News Texts" (May 2023). doi: 10.28995/2075-7182-2022-20-XX-XX.

[14] J. Tang, "Leveraging Transformer Neural Networks for Enhanced Sentiment Analysis on Online Platform Comments", 4th International Conference on Machine Learning and Computer Application (ICMLCA 2023), ACM, pp. 256–260 (Nov. 2023). doi: 10.1145/3650215.3650260.

[15] A. Sharma, and R. Bhalla, "Detecting Hate Speech for Hindi-English Code-Mix Text Data Using Dual Contrastive Learning", Sixth International Conference on Futuristic Trends in Networks and Computing Technologies (FTNCT06), vol. 259, pp 35–43 (2025). doi: 10.1016/j.procs.2025.03.304.

[16] A. Ilham, A. Bustamin, A. A. Prayogi, Safina, and A. R. Rahmika, "Development of Sentiment Analysis and Sarcasm Detection Systems for Social Network Users", 4th EPI International Conference on Science and Engineering, EICSE 2020 - Gowa, Indonesia, AIP Conference Proceedings, vol. 2543 (Nov. 2022). doi: 10.1063/5.0094737.

[17] https://www.kaggle.com/datasets/tegzes/isarcasm.

[18] T. Handoyo, Hidayaturrahman, C. J. Setiadi, and D. Suhartono, "Sarcasm Detection in Twitter - Performance Impact While Using Data Augmentation: Word Embeddings", *International Journal of Fuzzy Logic and Intelligent Systems,* vol. 22, no. 4, pp. 401–413 (Dec. 2022), doi: 10.5391/IJFIS.2022.22.4.401.

[19] N. Malali, "AI-Powered Data Preprocessing and Transformation Platform for Autonomous Data Cleaning, Advanced Fea", Patent no. 202521035175 (Apr. 2025).

[20] T. UcKan, "A hybrid model for extractive summarization: Leveraging graph entropy to improve large language model performance", *Ain Shams Engineering Journal,* vol. 16, no. 5, pp 1–15 (Mar. 2025). doi: 10.1016/j.asej.2025.103348.

[21] R. Setiyawan and Z. Mustofa, "Comparison of the performance of naive bayes and support vector machine in sirekap sentiment analysis with the lexicon-based approach", *Journal of Soft Computing Exploration,* vol. 5, no. 2, pp. 122–132 (June 2024), doi: 10.52465/joscex.v5i2.367.

[22] S. Adhikary, "A Machine Learning Approach for COVID-19 Tweet Classification", *Forum for Information Retrieval Evaluation, CEUR Workshop Proceedings*, vol. 3395, pp. 349–353 (2022).

[23] O. Mairaj, and S. U. R. Khan, "Unveiling temporal patterns in information for improved rumor detection", *Social Network Analysis and Mining,* vol. 15, no. 13, pp 1–14 (Mar. 2025). doi: 10.1007/s13278-025-01421-2.

Received January, 2025