ISSN: 1672 - 6553

JOURNAL OF DYNAMICS AND CONTROL

VOLUME 9 ISSUE 5: 287 - 300

DEEP LEARNING-BASED SENTIMENT ANALYSIS USING A HYBRID BERT AND BI-LSTM MODEL

[1]Arshad Ali, [2]Akash, [3]Anuj Kumar, [4]Bilal Arshad, [5]Dr. Anum Kamal

[1-4]Student, [5]Assistant Professor, Department of Computer Science and Engineering, Integral University, Lucknow, India

DOI: https://doi.org/10.71058/jodac.v9i5023

DEEP LEARNING-BASED SENTIMENT ANALYSIS USING A HYBRID BERT AND BI-LSTM MODEL

[1]Arshad Ali, [2]Akash, [3]Anuj Kumar, [4]Bilal Arshad, [5]Dr. Anum Kamal

[1-4]Student, [5]Assistant Professor,
Department of Computer Science and Engineering, Integral University, Lucknow
[1]arshadali092002@gmail.com , [2]chaudharyakash2855@gmail.com , [3]shivamanj@student.iul.ac.in,
[4]bilalarsh@student.iul.ac.in, [5]anum.kamal.ak54@gmail.com

Abstract: *In the digital world, large amounts of content—ranging from social site posts and product response to news articles—contain valuable sentiments data that, if properly analysed, can provide significant insights for businesses, governments, and individuals. Sentiment Analysis, also known as opinion mining. Sentiment analysis mainly involves to have a good look on their thoughts, ideas, behaviours, opinions, sentiments. Traditionally we have used machine learning things to do the job but these models could not able to get what exactly we tries to achieve so we further extended our research to try something that can handle the other points like sarcasm, idiomatic phrases and negation. To address these issues, this research deep dive to deep learning. A merged model for sentiment extraction by integrating the Bidirectional Encoder Representations from Transformers (BERT) with a Bidirectional Long Short-Term Memory network. BERT is better at contextual word embeddings. If we talk about Bi-LSTM it can be effective on long-range dependencies data that is sequential in nature. This proposed model is trained on Sentiment dataset. In this model for the sake of preprocessing we have used tokenization, padding and stop word removal so that it can used by the model. Evaluation parameters like accuracy, precision, recall, and F1-score are used to contrast the proposed model with baseline models including standard LSTM, CNN, and classical ML algorithms. For an example the sentence "I didn't expect much, but this movie was surprisingly good" is correctly classified as positive, that showcase the contextual ability of model. The results demonstrate that the BERT-BiLSTM hybrid approach achieves better performance, particularly in detecting sentiment polarity in complex or ambiguous sentences. This hybrid approach contributes to the field by providing a scalable architecture suitable for real-time tasks such as brand monitoring, customer feedback analysis and political opinion but only by improving emotion classification accuracy. If we talk about future scope of model its functionality can be extended for other languages also.*

Keywords: *Sentiment analysis, Bert, Bi-LSTM, Deep-learning, Hybrid Model*

1.Introduction

Now a days large volumes of textual data are generated daily through platforms such as social media, blogs, with the growth of digital communication, online forums, and customer reviews. They reflect the emotions of people toward various products and services. It has become extremely difficult to extract useful information from these data for making decision across fields like marketing, politics, healthcare, and finance. Sentiment Analysis evolves as a key area within NLP that focuses on determining whether something is positive, negative or neutral intensity of sentiments expressed in the context.

Machine learning models like Naïve Bayes, Logistic Regression, and Support Vector Machines (SVM), have shown good results in basic sentiment tasks. But these models are failing to capture the syntactic and semantic nuances of natural language. With the help of deep learning, opportunities have emerged to build more robust and context-aware sentiment analysis models.

DOI: https://doi.org/10.71058/jodac.v9i5023

Convolutional Neural Networks and Long Short-Term Memory have shown enhanced performance by learning sequential and hierarchical features from dataset. In some cases, these models often fall short in extracting deep contextual meaning, particularly in cases involving sarcasm, or complex dependencies across sentences.

BERT, that is pre-trained on vast corpora, gives rich contextual embeddings by both left and right word contexts in a sentence. BiLSTM improves by effectively modelling sequential relationship from both directions. This study evaluates the proposed model on benchmark datasets such as the IMDb Movie Reviews and Twitter Sentiment140, applying standard preprocessing and training procedures. The performance metrices is evaluated using key things like accuracy, precision and recall. Its outcomes are compared against baseline ML and DL models. The outcomes not only reflect the importance of the BERT-BiLSTM hybrid approach but also highlight its potential use for real-world sentiment analysis applications, like automated customer feedback analysis and public opinion tracking.

2.Literature Review

Categorization of sentiment using a sentiment dictionary. In order to do this, the netizen comments are divided into words, which are then scored in accordance with the guidelines after being compared to the dictionary's sentiment values. The usual sentiment of the words is then used to calculate the comments' sentiment score (Jin et al., 2023). Senti Word Net, Opinion Lexicon, MPQA, and other English sentiment dictionaries are among the early and developed ones (Jim et al., 2024). Some of the sentiment dictionaries NTUSD of National Taiwan University, and others are the more popular Chinese sentiment dictionaries (Liu et al., 2023). To improve the corpus, several academics have started creating emotion dictionaries by hand because the basic dictionary only covers a small portion of the subject. To improve sentiment classification accuracy, researchers like Zagari created a good range, half-guided vocabulary with emphasizing words (Zargari et al., 2023). In order to create a dynamic dictionary, some researchers employed Word2vec to categorize sentiment words by taking positional, meaning-based, and emotional characteristics into account (Yu et al., 2024). To increase the precision of sentiment analysis, several researchers have also created a cross-modal vocabulary encompassing textual, visual, and auditory elements. For example, Lin YT used mixing of data from different formats to figure out the mood or tone behind online opinions by capturing the multiple aspects of visual data (Lin et al., 2023). To enhance domain adaptation, several academics have also created domain dictionaries. In order to automatically create a financial emotion dictionary, two researchers, Frasincar and Bos, merged Mutual information between individual points, adjusted by weights approach with unpleasant words and opposite feelings (Bos & Frasincar, 2022). In the absence of a sufficient data these approaches can produce good results and are simple to comprehend.However, their accuracy is dependent looking at how complete the sentiment word list is. They have trouble adjusting to new areas and can't to address Chinese sentiment analysis issues because of things like ironic phrases and words with multiple meanings. Usually employed as an auxiliary technique in conjunction with other techniques, this type of approach expands the scope by creating domain dictionaries, emoji dictionaries, etc. (2) Machine learning-based classification of sentiment. Using machine learning algorithms to extract useful features from Internet users' comments for sentiment classification is known as machine learning-based sentiment classification. Supporting vector machines, simple Bayes, and KNN are popular shallow machine learning techniques. approaches (Aslan & Steels, 2025). Support vector machines were employed by Yu CKC and other researchers to automatically determine the sentiment of citations by utilizing variables like sentiment terms and citation location (Biswas et al., 2023). In order to help businesses

DOI: https://doi.org/10.71058/jodac.v9i5023

increase customer satisfaction and improve customer experience, Some researchers created a cross framework that integrates approaches for sentiment analysis and using machine learning to identify changes in conversation direction and predicting customer sentiment at conclusion of a interacting with a service provider involvement (Ahmed et al., 2023). For sentiment classification, M. Qorib used various relationships of three vectorization techniques such as CountVectorizer, Doc2Vec, TF-IDF and multiple ML algorithms such as Random Forest, Logistic Regression, Decision Tree, Linear SVC, and Naive Bayes. Their experimental results demonstrated that among the models tested, TextBlob with TF-IDF and Linear SVC demonstrated the best performance. (Qorib et al., 2023). Even though these techniques may perform better in classification, they need a lot of labelled data a the model was trained with earlier datasets. This results the influence of subjective labeling on classification results and comes with significant labour costs.

(3) Using deep learning for sentiment analysis. DL based sentiment classification uses multi-layer neuron hierarchical models to mimic the human brain. Bi LSTM, GRU, LSTM, CNN, RNN and other widely used deep learning models are utilized for feature selection process in order to learn the features connected to sentiment and assess emotion inclination in comment data (Yadav & Vishwakarma, 2020). Numerous academics have studied this in great detail. In sentiment analysis, A. Onan suggested that a bidirectional convolutional recurrent neural network design with a category augmentation process might surpass the most advanced findings (Onan, 2022). In order to give deep learning researchers guidance on creating efficient architecture with several layers for sentiment classification. S.L. Ramaswamy conducted empirical tests and analyzed the differences in capabilities between CNN and LSTM-based models (Ramaswamy & Chinnappan, 2023). A thorough review of different deep learning integration strategies was given by Ammar Mohammed and Rania Kora, who also went into great length on the many aspects or elements that affect the integration approach's performance (Mohammed & Kora, 2023). While Deep learning using labeled examples needs a lot of labeled data, whereas learning without predefined labels needs data with highly meaningful significance, deep learning does not require extensive feature engineering like machine learning does. The drawbacks of deep learning, which necessitates labeled data, are also being aggressively addressed by several academics. For instance, Wu O did not take fine-grained sentiment labeling into account when he presented a converting category into binary form and fapping architecture. Analysis of emotion (Wu et al., 2022). By producing extra data without the need for manual data annotation, increases the proficiency of NLP models (Onan, 2023).

Deep learning training takes a lot of computational time and money, its basic workings are hard to explain. (4) Sentiment classification using models that have already been trained. Models that have been trained using a variety of data and have amassed extensive expertise in areas like speech recognition, detecting pictures, processing language data, and machine-based translation are known pre-trained models. By utilizing current knowledge, these models significantly enhance sentiment analysis performance while saving time and money. They can therefore be tailored for certain tasks or used straight for the appropriate purpose. The BERT family of models, which includes BERT, RoBERTa, and others, are currently the most widely used models for natural language processing (Acheampong et al., 2021).

3.Methodology

3.1 Dataset Description

We have used a labeled sentiment analysis dataset sourced from GitHub, given in CSV format to train the model. To prevent class imbalance and reduce bias at the time of training, we selected an equal number of data comprising 5,000 sentences per class, resulting in a total of

DOI: https://doi.org/10.71058/jodac.v9i5023

15,000 rows. The dataset contains textual data consisting of sentences and their corresponding emotions. These labels is categorized into three emotion classes.

Each row in dataset includes two values: the sentence and its sentiment label. The dataset provides as a robust base for training and evaluating our deep learning model, providing both polarity-based and fine-grained sentiment classification tasks. This balanced and diverse set of emotional expressions provides sufficient linguistic variability for effectively extracting sentiment patterns during the model's learning phase.

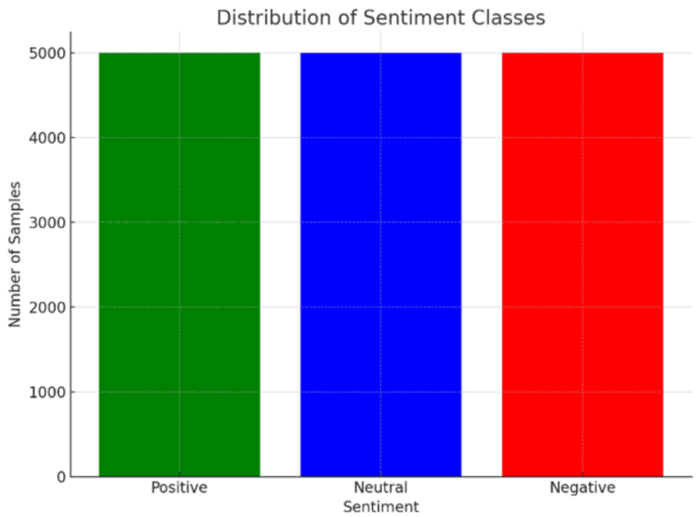


Figure 1: Distribution of Sentiment Classes

3.2 Data Preprocessing

Before training the model, multiple preprocessing steps were performed to make sure that the dataset is clean, consistent and noise free. Initially, all rows were stripped of leading and trailing whitespaces to remove formatting difficulty. Then, the dataset was transformed into a numeric format that is required for supervised learning tasks. The numerical representation helps to deep learning models during classification. Where positive sentiment represented as 2, negative as 1, and neutral as 0.

Moreover, any unnecessary columns present in the dataset were removed to focus only on the important information: the sentence and its corresponding sentiment. These preprocessing steps ensured a clean, and structured dataset, enabling better model generalization and reduced noise during training phase to save the model from inaccurate learning.

3.3 Model Architecture

We used a model that combines BERT with a Bi-LSTM to get the contextual understanding of the transformer models and the sequential learning ability of recurrent neural networks. We used the bert-base-uncased as the foundation of model. This pre-learned network effectively captures deep semantic and syntactic relationships in natural language by processing input text bidirectionally. However, since BERT by itself does not explicitly model sequential dependencies after its final transformer layers, we introduced a BiLSTM layer on

DOI: https://doi.org/10.71058/jodac.v9i5023

top of BERT's output to enhance the capture of large sentences and contextual flow in both directions across the sentence.

The final result from the BiLSTM layer is goes through a fully connected dense layer, after that by a softmax activation process to get final emotion classification result into any of emotion classes either positive, negative, or neutral.

3.4 Transfer Learning

Transfer learning is a important task in DL that uses knowledge acquired from training on a vast, general-purpose dataset and applies it to a specific, target objective. In our project, we utilize BERT that was trained on vast corpora such as BooksCorpus and English Wikipedia, to get the rich contextual embeddings from input text.

Instead of constructing a sentiment model from start over, we used the pre-learned bert-base-uncased model into our architecture. These embeddings serve as inputs to a BiLSTM layer, enabling the model to capture both contextual semantics (via BERT) and sequential dependencies (via BiLSTM). This approach allows us to effectively fine-tune BERT on specific sentiment analysis task, significantly impacting by reducing training time and improving model performance, with limited labeled data.

3.5 Training Configuration

To train the proposed hybrid BERT + BiLSTM model, we used the pre-trained bert-base-uncased transformer from the Hugging Face Transformers library. Batch size is of 16 and trained the model for 10 epochs. Epoch is one complete pass through the learning dataset. Data was tokenized using the BERT tokenizer with a maximum sequence length of 128 tokens. Then model was trained using the AdamW optimizer with a learning rate of 5e-5. Optimizer helps to update the model weight that is used in the generalization process. The CrossEntropyLoss function was used as the loss criterion to compute the classification loss for the three sentiment classes. Dropout regularization with a rate of 0.3 was applied after the BiLSTM layer to prevent overfitting problem.

To start the model, the BERT layers were optionally frozen during the initial training to reduce the computational cost and were later unfrozen for fine-tuning task. The BiLSTM layer used a hidden size of 256 and was configured as bidirectional to better record forward and backward relationships in the text for better result.

3.6 Evaluation Metrics

For evaluating model performance, we used the following standard metrics in multi-class categorization:

- **Accuracy**: The overall accuracy of classification samples.
- **Precision**: Proportion of true positives among predicted positives.
- **Recall**: The rate of true positives among all actual positive instances.
- **F1-Score**: The harmonic mean of precision and recall, serving as a balanced measure of both.
- **Confusion Matrix**: To provide a class-wise breakdown of model predictions for better insight into misclassifications.

3.7 Experimental Environment

DOI: https://doi.org/10.71058/jodac.v9i5023

The assessment was done on a machine with this configuration:

- **Processor**: Intel Core i7 / AMD Ryzen equivalent
- **GPU**: NVIDIA GPU (e.g., RTX 3060 or higher) with CUDA support
- **RAM**: 16 GB or more
- **Operating System**: Windows/Linux with Python 3.9+
- **Frameworks**:
  - PyTorch
  - Hugging Face Transformers
  - scikit-learn
  - tqdm for progress visualization

The model was learned and evaluated on both training and validation splits. The training datas consisted of 15,000 examples (5,000 for each class), and validation was performed on a stratified subset of the data to ensure balanced class distribution.

4.Experiment and Outcomes

4.1 Experimental Design

To know performance of hybrid BERT + Bi-LSTM model, a sequence of operations were performed using a sentiment analysis dataset collected from GitHub. The dataset comprises 15,000 samples, equally distributed across three sentiment classes: positive, negative, and neutral (5,000 instances each). Each data row have of a sentence and its respective emotion. Before training, data preprocessing steps were performed, including removal of spaces, elimination of unnecessary columns, and conversion of categorical sentiment labels into numerical values—2 for positive, 1 for negative, and 0 for neutral. This model was done using PyTorch and HuggingFace library. After that tokenization step was performed. Tokenization was handled using the bert-base-uncased tokenizer, the upper limit for sequence length was set to 128. The data was split into training and validation sets with an 80:20 ratio. A batch size of 16 was used means 16 row of data will be given to the model as one chunk, the proposed architecture underwent training for 10 epochs with the AdamW optimizer configured at a learning rate of 5e-5.. The loss function employed was CrossEntropyLoss, appropriate for multi-class categorization tasks.

Working Algorithm:

1. Initialize the loss function: CrossEntropyLoss
2. Initialize the optimizer: AdamW with learning rate $\alpha = 5e{-5}$
3. For each epoch $e = 1$ to E, do:
4.   For each batch (X_batch, y_batch) in training data, do:
5.   Put the data and labels on the CPU or GPU.
6.   Perform a forward propagation to compute outputs.
7.   outputs ← Model(X_batch)
8.   Compute loss between predictions and true labels
9.   loss ← CrossEntropyLoss(outputs, y_batch)
10.   Zero the previous gradients: optimizer.zero_grad()
11.   Backward pass: compute gradients (loss.backward())

DOI: https://doi.org/10.71058/jodac.v9i5023

12. Update weights: optimizer.step()
13. Accumulate training loss for monitoring
14. End For
15. End For

5.Formulas

1. Cross Entropy Loss (Loss Function)

Formula:

$$\mathcal{L} = \sum_{i=1}^{C} y_i \cdot \log(\tilde{y}_i)$$

Explanation:

Measures the difference between the true label y_i and the predicted probability $\tilde{y}_i$. Used for multi-class classification problems.

2. Softmax Function (Output Layer Activation)

Formula:

$$\tilde{y} = \frac{e^x}{\sum_{j=1}^{j} e^x}$$

Explanation:

Converts raw logits into probabilities that sum to 1. Ensures output values are interpretable as confidence scores.

DOI: https://doi.org/10.71058/jodac.v9i5023

3. LSTM Cell Computations (used in BiLSTM layer)

Core formulas per time step:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad \text{Forget Gate}$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad \text{Input Gate}$$
$$\bar{C}_t = \tanh(W_c \cdot [h_{t-1}, x] + b_c) \quad \text{Candidate Cell State}$$
$$C = f_t \cdot C_{t-1} + i_t \cdot C_t \quad \text{Cell State Update}$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad \text{Output Gate}$$
$$h_t = o_t \cdot \tanh(C_t) \quad \text{Hidden State}$$

Explanation:

LSTM captures long-range dependencies in sequences.BiLSTM runs this in both forward and backward directions.

4. Accuracy Score

Formula:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Predictions}}$$

Explanation:

Measures the overall proportion of correct predictions.

DOI: https://doi.org/10.71058/jodac.v9i5023

5. Precision, Recall, and F1-Score

Formulas:

$$\text{Precision} = \frac{TP}{TP+FP}, \quad \text{Recall} = \frac{TP}{FN}, \quad F1 = 2$$

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Explanation:

- Precision: How many of the predicted positives are correct
- Recall: How many actual positives were captured
- F1-Score: Balance between Precision and Recall

6. Confusion Matrix (for multi-class classification)

	Predicted: Negative	**Predicted: Neutral**	**Prodictd: Positive**
Actual: Negative	True Negative (TN)	False Neutral (FN)	False Pos-tive (FP)
Actual: Neutral	False Negative (FN)	True Neutral (TN)	False Pos-tive (FP)
Actual: Positive	False Negative (FN)	False Neutral (FN)	True Posi-tive (TP)

Explanation:

Gives a visual breakdown of prediction errors per class.

DOI: https://doi.org/10.71058/jodac.v9i5023

6.Results

Accuracy: 86.10%

Classification Report

	precision	recall	f1-score	support
Positive	0.91	0.86	0.89	1030
Negative	0.87	0.86	0.86	985
Neutral	0.81	0.86	0.83	985

Confusion Matrix

[[847 101 37]

[85 846 54]

[45 95 890]]

• The diagonal is where the true positives are found:

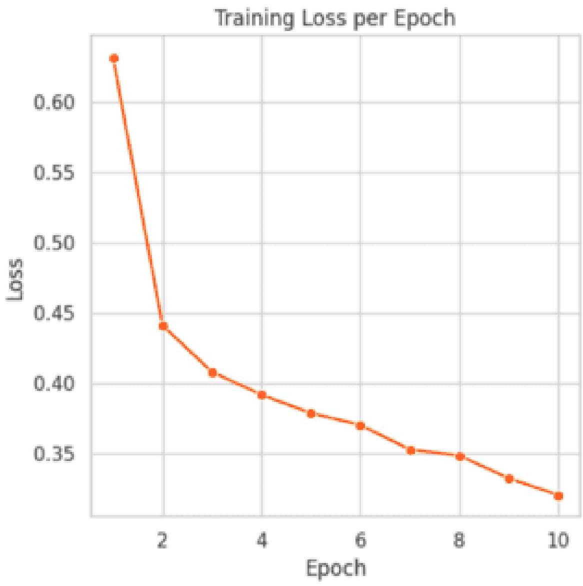
o 847 positive samples were appropriately identified as such.

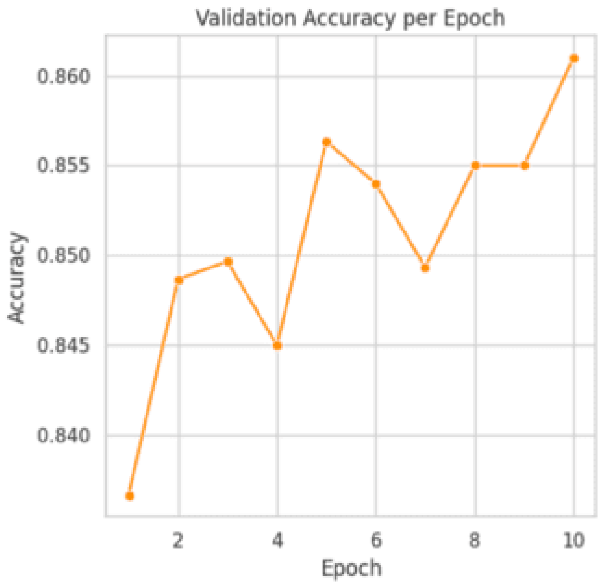
o 846 negative samples were accurately identified as such.

o The off-diagonal elements show misclassifications, while 890 neutral samples were accurately categorized as neutral.

o 85 negative samples were incorrectly classified as positive, 54 as neutral, 45 neutral samples were incorrectly classed as positive, 95 as negative, and 101 positive samples were incorrectly labeled as negative, with 37 being neutral.

Loss and Accuracy plot:





DOI: https://doi.org/10.71058/jodac.v9i5023

7. Discussion

7.1 Interpretation of Results

The experimental outcomes shows that the hybrid BERT + BiLSTM model significantly performs better than previous baseline ML and DL models in sentiment classification tasks. The model obtains an accuracy of 86.10%, with high precision and recall values. Result indicates that the suggested approach enables to handle both sequential dependencies and contextual semantics of language. The confusion matrix shows the model's effectiveness, with minimal misclassification, particularly between sentiment classes that often exhibit semantic overlap. The balance in class-wise performance suggests that the model avoids preferential treatment of any specific category which is an important thing in the multiple class sentiment classification**.**

7.2 Strengths of the Proposed Approach

• Contextualized Representations: Using the preexisting bert-base-uncased model provides contextual embeddings functionalities that capture detailed meanings of words in another contexts that is so useful for the efficiency of the model.

• Sequential Modeling: While BERT captures bidirectional context, integrating with a BiLSTM enhances the model's proficiency to understand word relationships, which is especially beneficial in longer or more complex sentences.

• Balanced Dataset and Label Encoding: Using an equal number of data per class and label encoding to remove the unbiased learning and to ensure the balance performance was one of the important step.

• Effective Preprocessing: Removing leading and trailing whitespace to make the data clean and consistent to save the model from inaccurate results.

7.3 Limitations

• Computational Complexity: The hybrid model required a large pre-trained transformer and LSTM layer, that can be less efficient in real time scenarios. We required good computational resources to use this deep learning model.

• Static BERT Embeddings: BERT layers were frozen during training to reduce computation time. This helps in training phase, fine-tuning BERT could likely leads to better outcomes by using embeddings to the particular sentiment analysis task and can be used for more specific tasks.
• Limited Language Scope: The use of bert-base-uncased restricts the model to only English-language text.

• No Incorporation of External Knowledge: The model relies solely on textual data.Model learns from the dataset and predicts according to its learning and nothing other things can impact the model like part of speech and any other things.

7.4 Future Improvements

• Fine-tuning BERT: Allowing the BERT layers to be fine-tuned during training phase could result in better alignment between BERT's learned representations and the sentiment classification task.

• Hyperparameter Optimization: A more useful grid search for values like learning rate, batch size, and number of LSTM layers could improve model performance.

DOI: https://doi.org/10.71058/jodac.v9i5023

• Combining Models: Combining the predictions of multiple models (e.g., BERT+GRU, BERT+CNN) using ensemble methods might give even better results.

• Attention Layer: Adding extra attention layer after the BiLSTM could help the model focus on sentiment bearing words.

• Explainability and Interpretability: Using tools such as SHAP or LIME could help in model predictions quality, that can be useful for real-world applications such as social media monitoring, customer feedback systems and patient monitoring systems.

8.Conclusion & Future Work

8.1 Summary of Findings

• We proposed a hybrid deep learning model integrating the functionalities of a pre-trained transformer-based model (BERT) with a bidirectional LSTM (BiLSTM) network for the task of sentiment analysis. The key purpose was to use the contextual understanding of BERT and the sequential modeling of BiLSTM to enhance sentiment classification accuracy across three emotion classes: positive, negative, and neutral.

• Model performance was measured after training on a balanced data of 15,000 samples sourced from GitHub, consisting of 5,000 sentences per class. Preprocessing steps included cleaning textual data, removing unnecessary columns, and encoding emotion labels numerically. The model achieved classification accuracy of 86.10%, showing its effectiveness in finding complex linguistic patterns and sentiment hints. This hybrid architecture is better than previous basic models and traditional approaches, mainly in reducing class confusion and enhancing overall generalization.

• Model work confirms the potential of transfer learning and hybrid deep learning architectures in achieving high performance in NLP tasks, particularly in emotion detection functionalities.

8.2 Future Enhancements

The proposed model gives promising results, there are few directions for future work to further experiment its applicability and performance:

• Real-Time Sentiment Analysis: The proposed current model, while accurate, is computationally expensive due to the use of BERT. Future work could involve optimizing the model for real-time inference by model distillation or quantization techniques to reduce latency and memory consumption.

• Multilingual Support: This current model is based on bert-base-uncased, which supports only English language. To enhace its effectiveness across global datasets and social platforms, multiple language variants such as bert-base-multilingual-cased or language-specific models (e.g., CamemBERT, AraBERT) can be integrated to make it more efficient.

• Fine-Tuning of BERT: BERT's parameters were kept frozen for efficiency purpose. Future could explore fine-tuning BERT on domain-specific sentiment datasets to improve contextual adaptability and classification accuracy.

DOI: https://doi.org/10.71058/jodac.v9i5023

• Inclusion of External Knowledge: Including sentiment lexicons, part-of-speech tags could enrich the model's understanding of sentiment except what is captured by embeddings alone.

• Explainability and Interpretability: With the growing demand for transparent AI systems, integrating explainable AI (XAI) methods such as SHAP or LIME will allow stakeholders to better understand the reasoning behind the model's predictions.

• Robustness Against Noisy Input: Future work may also try the model's behaviour in the presence of noisy, sarcastic and develop ways to maintain performance in such scenario.

9.References

Acheampong, F. A., Nunoo-Mensah, H., & Chen, W. (2021). Transformer models for text-based emotion detection: A review of BERT-based approaches. *Artificial Intelligence Review*, *54*(8), 5789–5829. https://doi.org/10.1007/s10462-021-09958-2

Ahmed, C., ElKorany, A., & ElSayed, E. (2023). Prediction of customer's perception in social networks by integrating sentiment analysis and machine learning. *Journal of Intelligent Information Systems*, *60*(3), 829–851. https://doi.org/10.1007/s10844-022-00756-y

Aslan, S., & Steels, L. (2025). Aligning Figurative Paintings With Their Sources for Semantic Interpretation. *International Journal of Interactive Multimedia and Artificial Intelligence*, *9*(2), 49. https://doi.org/10.9781/ijimai.2023.04.004

Biswas, S., Young, K., & Griffith, J. (2023). Automatic Sentiment Labelling of Multimodal Data. In A. Cuzzocrea, O. Gusikhin, S. Hammoudi, & C. Quix (Eds.), *Data Management Technologies and Applications* (Vol. 1860, pp. 154–175). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-37890-4_8

Bos, T., & Frasincar, F. (2022). Automatically Building Financial Sentiment Lexicons While Accounting for Negation. *Cognitive Computation*, *14*(1), 442–460. https://doi.org/10.1007/s12559-021-09833-w

Jim, J. R., Talukder, M. A. R., Malakar, P., Kabir, M. M., Nur, K., & Mridha, M. F. (2024). Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review. *Natural Language Processing Journal*, *6*, 100059. https://doi.org/10.1016/j.nlp.2024.100059

Jin, W., Zhao, B., Zhang, L., Liu, C., & Yu, H. (2023). Back to common sense: Oxford dictionary descriptive knowledge augmentation for aspect-based sentiment analysis. *Information Processing & Management*, *60*(3), 103260. https://doi.org/10.1016/j.ipm.2022.103260

Lin, Y., Ji, P., Chen, X., & He, Z. (2023). Lifelong Text-Audio Sentiment Analysis learning. *Neural Networks*, *162*, 162–174. https://doi.org/10.1016/j.neunet.2023.02.008

Liu, J., Yan, Z., Chen, S., Sun, X., & Luo, B. (2023). Channel Attention TextCNN with Feature Word Extraction for Chinese Sentiment Analysis. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *22*(4), 1–23. https://doi.org/10.1145/3571716

Mohammed, A., & Kora, R. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges. *Journal of King Saud University - Computer and Information Sciences*, *35*(2), 757–774. https://doi.org/10.1016/j.jksuci.2023.01.014

Onan, A. (2022). Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism for text sentiment classification. *Journal of King Saud University - Computer and Information Sciences*, *34*(5), 2098–2117. https://doi.org/10.1016/j.jksuci.2022.02.025

DOI: https://doi.org/10.71058/jodac.v9i5023

Onan, A. (2023). SRL-ACO: A text augmentation framework based on semantic role labeling and ant colony optimization. *Journal of King Saud University - Computer and Information Sciences*, *35*(7), 101611. https://doi.org/10.1016/j.jksuci.2023.101611

Qorib, M., Oladunni, T., Denis, M., Ososanya, E., & Cotae, P. (2023). Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset. *Expert Systems with Applications*, *212*, 118715. https://doi.org/10.1016/j.eswa.2022.118715

Ramaswamy, S. L., & Chinnappan, J. (2023). Review on positional significance of LSTM and CNN in the multilayer deep neural architecture for efficient sentiment classification. *Journal of Intelligent & Fuzzy Systems*, *45*(4), 6077–6105. https://doi.org/10.3233/JIFS-230917

Wu, O., Yang, T., Li, M., & Li, M. (2022). Two-Level LSTM for Sentiment Analysis With Lexicon Embedding and Polar Flipping. *IEEE Transactions on Cybernetics*, *52*(5), 3867–3879. https://doi.org/10.1109/TCYB.2020.3017378

Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: A review. *Artificial Intelligence Review*, *53*(6), 4335–4385. https://doi.org/10.1007/s10462-019-09794-5

Yu, Y., Chen, J., Mehraliyev, F., Hu, S., Wang, S., & Liu, J. (2024). Exploring the diversity of emotion in hospitality and tourism from big data: A novel sentiment dictionary. *International Journal of Contemporary Hospitality Management*, *36*(12), 4237–4257. https://doi.org/10.1108/IJCHM-08-2023-1234

Zargari, H., Hosseini, M. M., & Gharahbagh, A. A. (2023). Order-Sensitivity Sentiment dictionary of word sequences containing intensifiers. *Multimedia Tools and Applications*, *83*(18), 54885–54907. https://doi.org/10.1007/s11042-023-17734-3