RESEARCH ARTICLE

Boosting Viewer Experience with Emotion-Driven Video Analysis: A BERT-based Framework for Social Media Content

Zubair Akbar[1,*], Muhammad Usman Ghani[2] and Umair Aziz[2]

[1] Department of Computer Science, Air University, Multan 60000, Pakistan
[2] Department of Computer Science, National College of Business Administration & Economics, Multan 60000, Pakistan

Abstract

social media has significantly transformed the digital landscape by enabling an unprecedented expansion of content, further accelerated by the COVID-19 pandemic, which increased the demand for online classes, virtual meetings, and recorded conferences. While major technology companies have previously employed sentiment analysis and opinion mining to gauge user feedback, this study proposes a novel framework for emotion-based video content analysis. The proposed method extracts audio from social media videos and applies Speech-to-Text (STT) conversion. The extracted text is then processed using a pre-trained BERT model, leveraging its fine-tuned capabilities and 110 million parameters to enhance emotion recognition. To improve inference, we modify the initial embedding layers of BERT to refine emotional analysis for unseen video content, ensuring better alignment with viewers' emotional responses. Experimental results demonstrate that the pre-trained BERT model outperforms traditional deep learning and machine learning approaches, achieving 83% accuracy and an F1 score of 83. Comparatively, CNN and LSTM models achieved 74% and 73% accuracy, respectively, while SVM resulted in 72% accuracy. The proposed framework offers a more refined emotional analysis, potentially improving user engagement by making content more relatable and emotionally intuitive for viewers.

1 Introduction

Every day, a significant amount of video content is uploaded to social media platforms, including medical and health-related videos. These videos play a key role in educating patients, assisting in disease diagnosis, and supporting medical research. The rapid increase in video consumption, especially during the COVID-19 pandemic, highlights the growing role of video in online health consultations, virtual medical meetings, and educational conferences. For instance, YouTube saw its monthly video views grow from 4.67 billion

to 11 billion due to the pandemic, with substantial contributions from online health-related content. This tremendous expansion of video content has had a significant impact on healthcare and bioinformatics, particularly in the extraction and analysis of emotions expressed in videos [1]. Emotion analysis in video content, especially in medical consultations or health-related discussions, can provide valuable insights for healthcare providers.

Due to the massive growth in video content on social media, the extraction of videos in keeping with emotions is a burdensome task for the user. In the past, online media sentiment research was mainly to calculate the sentimentality of online media. Emotion and sentiment analysis have become a new trend in social media, helping users and companies to automatically extract user-generated content, especially opinions expressed in videos and recommendation systems [2]. Emotions and sentiment are different from each other but sentiment relies on emotions while emotions are fleeting states of mind that come and go. Sentiments, on the other hand, last longer in our bodies and minds if we take the opportunity to learn about them [3]. Emotions occur unconsciously and will emerge and disappear quickly. When our mind intervenes and we recognize what we're doing, we call it a feeling. It takes time for sentiments to emerge from emotions. Emotions are more powerful as a result of this; sentiments necessitate the activation of more dynamic systems. While the Sentiment is often preceded by emotion. We can't have sentiments if we don't have impulses. In the same individual, the same emotion will elicit a variety of responses [4].

The basic method was to use natural language processing, text analysis, and computational linguistics to dig out people's opinions, sentiments, evaluations, attitudes, and attitudes from the text. The basic deployment is to build a system through some knowledge bases, plus some basic principles of statistics, to classify the text of the network, and to get its polarity and polarity strength which is not perfect for dealing with the emotions of video content for the user [5].

However, the traditional machine learning algorithms mostly use the bag-of-words model for the representation, they also face problems such as sparse data features and the inability to extract emotional information [6]. The deep learning methods that have emerged in recent years can

well make up for the shortcomings of traditional machine learning methods. Deep learning models are represented by convolutional neural networks (CNN) and recurrent neural networks (RNN) while the objective of this work is to design a new framework that keeps the emotional information of video data for their viewers using the Bidirectional Encoder Representations model [7].

2 Related Work

Different researchers performed sentiment analysis for the judgment of opinion using different well-known techniques to realize the reviews of their users about their content and product [8]. According to the method of sentiment classification, the application research of sentiment analysis in the field of online product reviews can be divided into two categories. The machine learning method is mainly to learn the emotional characteristics of the training set, estimate the dependency between the input and output of the system, and then apply it to the classification judgment of the test set [9]. But the emotions and sentiment in both terminologies are different from each other, sentiment relies on emotions, and the emotions are fleeting states of mind that come and go. Sentiments, on the other hand, last longer in our bodies and minds if we take the opportunity to learn about them [10]. Emotions occur unconsciously and will emerge and disappear quickly. When our mind intervenes and we recognize what we're doing, we call it a feeling. It takes time for sentiments to emerge from emotions. Emotions are more powerful as a result of this; sentiments necessitate the activation of more dynamic systems [11].

Social networks are developing into an ecological platform that "connects everything" and especially video content is increasing exponentially. Due to this massive growth in video content on social media, the extraction of videos in keeping with emotions is a burdensome task for the user. In the past, online media sentiment research was mainly to calculate the sentimentality of online media [12]. The basic method was to use natural language processing, text analysis, and computational linguistics to dig out people's opinions, sentiments, evaluations, attitudes, and attitudes from the text. The basic deployment is to build a system through some knowledge bases, plus some basic principles of statistics, to classify the text of the network, and to get its polarity and polarity strength which is not perfect for dealing with the emotions of video content for the user [11, 13].

Existing literature has found a large number of techniques that can be used for multiple tasks in sentiment analysis, including supervised and unsupervised methods. Among the supervised methods, early papers used all supervised machine learning methods (such as support vector machines, maximum entropy, naive Bayes, etc.) and feature combinations [14]. Unsupervised methods include different methods using sentiment dictionaries, grammatical analysis, and syntactic patterns. There are many review books and papers, covering the early methods and applications extensively. All of these studies are based on traditional big giants to realize the reviews of the users about their content [15]. But none of these classify the emotions of social media videos from the perspective of a user and make users more relaxed by the selection of video as per their mode interest [16, 17].

3 Methodology

This section mainly focuses on the proposed framework to achieve the objective of the research. In this proposed framework, the social media video is considered as the primary research object, and different methods on it to classify the emotions of their users as shown in Figure 1.

The proposed framework consists of two subsystems that are performed the evaluate the emotions of social media video data using bidirectional encoder representations.

3.1 Training Pipeline

In the training pipeline of a purposed framework, there are two following sub-steps.

3.1.1 Preprocessing of Raw Data

First of all, the data needs to be preprocessed for the enhancement of the data quality to ultimately achieve the best results because it affects the overall performance of the system. In preprocessing, different methods as mentioned here.

Remove Special Character: The dataset contains grammatical words and other supporting words or POS to complete sentences that don't contain any core information about emotions. To increase the processing speed, remove the special characters from the dataset at the start of the procedure. With the help of Regular expression("regex"), the method removes all the special characters from the dataset and makes it more appropriate for further operations.

Case Conversion Operation: The conversion of the case utilizes the built function of python provided by the string library that converts all the upper case textual into the lower case to enhance the smoothness of the dataset.

Spell Correction: In spelling correction, correcting words of a textual dataset. There are many approaches but the adopted one is isolated words compute a list of spelling suggestions ranked by edit-distance, letter-n-gram similarity, or comparable measures.

Tokenization: Word Tokenization is the most commonly used tokenization algorithm. It splits a piece of text into individual words based on a certain delimiter. Depending upon delimiters, different word-level tokens are formed.

After the preprocessing operation, extracted the normalized data for the training of the model.

3.1.2 Bidirectional Encoder Representations Model

BERT (Bidirectional Encoder Representations from Transformers) has significantly advanced natural language processing (NLP) by introducing deep bidirectional contextual representations, enabling it to dynamically adapt to downstream tasks with task-specific modifications. Unlike earlier models such as ELMo, which rely on independently trained forward and backward LSTM cascades, BERT employs a more sophisticated bidirectional Transformer encoder, capturing dependencies in both directions simultaneously. This bidirectional learning mechanism enhances its ability to understand complex linguistic patterns, making it highly effective for tasks such as sentiment classification, emotion analysis, and text comprehension.

The pre-training phase of BERT utilizes two key objectives—Masked Language Modeling (MLM) and Next Sentence Prediction (NSP)—to develop rich contextual representations. MLM randomly masks input tokens, requiring the model to predict the missing words based on surrounding context, thereby improving its word-level semantic understanding. NSP, on the other hand, enhances sentence-level comprehension by training the model to determine whether two given sentences are contextually related. These techniques enable BERT to generate highly informative vector representations of words and sentences, ensuring optimal parameter initialization for subsequent fine-tuning tasks.

A critical component of BERT's superior performance lies in its input representation, which integrates token
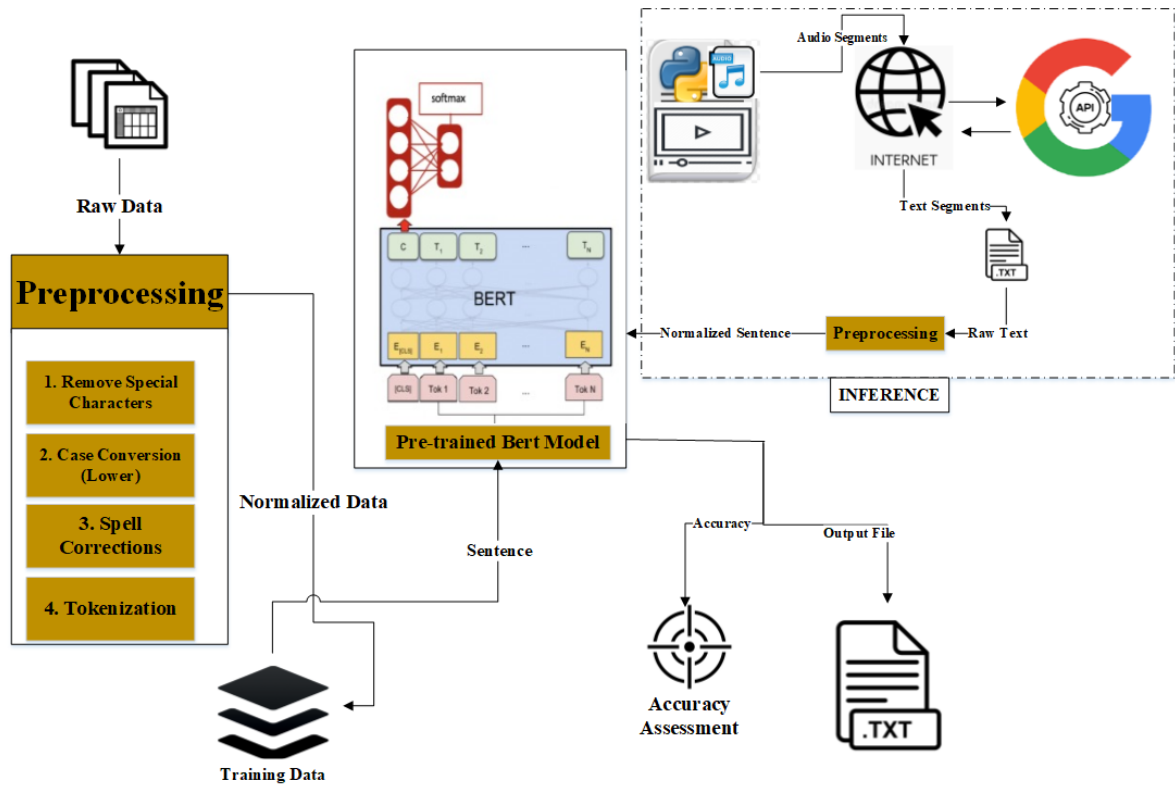
Figure 1. Emotion's evaluation framework for social media data.

embeddings (capturing individual words or sub words), segment embeddings (differentiating between multiple sentences), and position embeddings (encoding sequential information). The summation of these embeddings creates a rich, high-dimensional representation, reinforcing the model's ability to effectively capture linguistic context across different tasks. Due to its extensive parameterization and fine-tuning capabilities, BERT consistently achieves state-of-the-art performance in various NLP applications, including emotion recognition from text.

To ensure optimal performance in fine-tuning BERT for emotion analysis, a comprehensive hyperparameter tuning process was conducted, leveraging structured search strategies. The AdamW optimizer with weight decay was employed to improve optimization stability, while a linear learning rate decay with warm-up was utilized to prevent overshooting in early training stages. Several training configurations were tested, with experiments spanning 3, 5, 10, and 15 epochs to assess convergence and overfitting trends, ultimately determining that 5 epochs provided the best balance between generalization and computational efficiency.

To refine the model further, Bayesian optimization was employed to efficiently explore the hyperparameter space, outperforming traditional grid search and random search methods by prioritizing high-performing configurations. Key hyperparameters were systematically varied, including batch size (8, 16, 32, 64), learning rate (1e-5, 2e-5, 3e-5, 5e-5), dropout rate (0.1, 0.2, 0.3, 0.4), and maximum sequence length (128, 256, 512). The optimal settings batch size of 32, learning rate of 3e-5, dropout rate of 0.1, and sequence length of 256 were selected to enhance training stability and generalization while preventing overfitting. Additionally, gradient clipping was applied to mitigate exploding gradients, ensuring robust learning dynamics.

By leveraging BERT's bidirectional Transformer encoder and systematically optimizing hyperparameters, this framework delivers a highly effective solution for emotion-based analysis of video content. The integration of contextualized embeddings, structured learning rate schedules, and Bayesian hyperparameter tuning enhances the model's ability to accurately capture emotional nuances within textual data. These advancements position the proposed approach as a state-of-the-art methodology for sentiment and emotion classification, demonstrating its efficacy in analyzing user responses to social media and video-based content [12].

3.2 Inference Engine

Inference Engine is a core component that implements knowledge-based reasoning in an expert system. It is the realization of knowledge-based reasoning in a computer. It mainly includes two aspects of reasoning and control. It is an indispensable and important part of the knowledge system. The inference system compromises on the series of subsystems that incorporate to classify the emotions of the social media video from the viewer's perspective.

3.2.1 Speech to Text Sub System

Video to Audio Conversion: First of all, a specified video is selected from a social media platform to convert into audio by using the nonlinear editing algorithm of Python. In this way, wav file is generated for the specified video.

Audio chunks to text Segments Conversion: Now the duration of the audio wave file is calculated and the whole duration of the audio is divided into 5 seconds. In this way, every 5-second audio chunk is converted into text using Google API.

Preprocessing approach: Polish the textual data by removing the unnecessary data from the extracted segments of text to enhance the data quality for the Fine Tune Bert model. In preprocessing, we performed different methods as earlier mentioned.

3.3 Output File

By doing immense research, we have extracted the intensities of emotions by the model and stored it in a separate file that describes the emotions of the social media video.

Furthermore, we applied a "maximum algorithm" to extract the most dominant emotions and mapped them to social media-selected videos for the perspectives of viewers and boosted their number because of their phraseology to make him/herself more relaxed.

4 Experiments

To analyze the distribution of emotions in the dataset, we conducted an emotional word frequency analysis based on word reviews from all experimental text data. A tag cloud was generated, as shown in Figure 2, to visually represent the most frequently occurring emotional words. A tag cloud provides an intuitive depiction of text content by altering the font size and color of words based on their frequency, thereby highlighting the most prominent emotions expressed in the dataset. Words that

appear more frequently in the dataset are displayed in larger and bolder fonts, while less frequent words appear smaller, allowing for quick identification of dominant emotional expressions. The tag cloud serves as an effective summary tool for analyzing the semantic structure of the dataset and understanding the prevalence of different emotions.

Figure 2. Emotions word tag cloud.

The chosen dataset deals with the five essential emotions; sadness, joy, neutrality, anger, and fear. Each sentence of training data holds one or many emotions as shown in Figure 3, which indicates the distribution of the training data is unbalanced. Emotions of sadness joy and neutrality appear more commonly as compared to anger and fear. Meanwhile, negative examples are fewer as compared to positive examples for each type of emotion.

Figure 3. The emotional intensity distribution of text sentences in training data.

We applied four different algorithms (Pre-trained Bert, CNN, LSTM, and SVM) to extract the emotions from the video for their viewers. To evaluate the predictive ability of the model that identifies the best model among them we used Model Performance Evaluation criteria and calculated the Accuracy, Error Rate, Precision Rate, Recall Rate, and F1 Measurement.

4.1 Accuracy

Accuracy refers to the proportion of the number of correctly classified samples to the total number of samples. It is a commonly used method to evaluate the generalization ability of a learning model as shown in Table 1. It can be measured by equation (1).

$$\text{Accuracy} = \frac{n_{\text{True-Positives}} + m_{\text{True-Negatives}}}{N} \quad (1)$$

where $n_{\text{True-Positives}}$ and $m_{\text{True-Negatives}}$ are the number of correctly classified samples. N is the total number of samples. Both of these values can be extracted from the confusion matrix, which is the $M \times M$ matrix, where M is the number of predicted classes.

Table 1. Accuracy comparison.

Methods	Accuracy
CNN	74%
LSTM	73%
Pre Trained BERT	**83%**
SVM	72%

4.2 Error Rate

Error rate refers to the proportion of misclassified samples to the total number of samples as shown in Table 2 for selective algorithms [18]. It can be calculated using equation (2).

$$\text{Error Rate} = \frac{n_{\text{True-Negatives}} + m_{\text{False-Positives}}}{N} \quad (2)$$

where $n_{\text{True-Negatives}}$ and $m_{\text{False-Positives}}$ are the total number of wrongly classified samples. N is the total number of samples.

Table 2. Error Rate comparison.

Methods	Error Rate
CNN	0.26
LSTM	0.27
Pre-Trained BERT	**0.17**
SVM	0.28

4.3 Precision Rate

Precision Rate refers to the proportion of the true positive class among all the results predicted as the positive class as shown in Table 3. It can be defined as equation (3).

$$\text{Precision Rate} = \frac{n_{\text{True-Positives}}}{n_{\text{True-Positives}} + m_{\text{False-Positives}}} \quad (3)$$

where $n_{\text{True-Positives}}$ is the total number of actual classified samples, and $m_{\text{True-Positives}}$ is the total number of the wrong sample that is mistakenly classified to be correct.

Table 3. Precision comparison.

Methods	Precision Rate
CNN	0.75
LSTM	0.73
Pre Trained BERT	**0.83**
SVM	0.72

4.4 Recall Rate

Recall rate refers to the proportion of the number of correctly classified samples by the model to the number of all positive samples. While Table 4 showed the recall rate of all selected algorithms and using equation (4) calculate the value of recall rate of an algorithm.

$$\text{Recall Rate} = \frac{n_{\text{True-Positives}}}{n_{\text{True-Positives}} + m_{\text{False-Negatives}}} \quad (4)$$

where $n_{\text{True-Positives}}$ are the number of correctly classified samples, and $m_{\text{False-Negatives}}$ are the number of samples that are classified to be false.

Table 4. Recall Rate comparison.

Methods	Recall Rate
CNN	0.75
LSTM	0.73
Pre Trained BERT	**0.83**
SVM	0.73

4.5 F1 Measurement

F1-Measurement is the evaluation standard which is the harmonic average of the precision and recall and Table 5 shown the value of the F1 measure of selective algorithms. It can be measured by the given equation (5).

$$\text{F1 Score} = R = \frac{2 \cdot PR}{P + R} \quad (5)$$

where $P = $ Precision Rate and $R = $ Recall Rate.

Table 5. F1 Score comparison.

Methods	F1 Score
CNN	74
LSTM	73
Pre-Trained BERT	**83**
SVM	72

Evaluate the model's predictive ability on the test set according to the previously determined evaluation criteria and reveal that the Pre-Trained BERT model was found to be best as compared to all other models for this research work.

To explore the emotional, this research work takes sentence-level emotions as the research object which is taken from the SST (Speech to Text) operation of the extracted audio chunk of social media video. A sentence may contain a variety of emotions with different intensities and the Pre-trained BERT model is selected to dig out the intensity of emotion from the sentence-level text as shown in Figure 4.

Emotions	Intensities
Anger	2
Fear	6
Joy	33
Neutral	20
Sadness	2

Figure 4. Emotion intensities.

Furthermore, we applied a "maximum algorithm" to extract the most dominant emotions and mapped them to social media-selected videos for the perspectives of viewers and boosted their number because of their phraseology to make him/herself more relaxed.

4.6 Statistically Significant Test

To further assess the effectiveness of the pre-trained BERT model in comparison to traditional deep learning and machine learning approaches for emotion analysis, a paired t-test was conducted to evaluate whether the observed performance differences were statistically significant. This test determines whether the mean accuracy differences across multiple experimental runs (N = 10) arise from random variation or represent a genuine improvement in classification performance. The statistical evaluation included comparisons between BERT, CNN, LSTM, and SVM, with accuracy scores recorded over 10 independent runs for each model. The null hypothesis (H_o) posited no significant difference in mean accuracy between BERT and the baseline models, while the alternative hypothesis (H_1) assumed that BERT outperforms the baselines. The results, presented in Table 6, indicate that BERT achieved significantly higher accuracy than CNN (+9%), LSTM (+10%), and SVM (+11%), with t-values of 5.12, 4.87, and 6.32, respectively. The p-values for all comparisons (p < 0.05) confirm the statistical significance of these differences, reinforcing that the superior performance of BERT is unlikely to be attributed to chance. These findings underscore the robust advantage of BERT's bidirectional Transformer-based architecture, which captures deep contextual dependencies more effectively than the sequential processing mechanisms in CNN and LSTM or the feature-based approach in SVM. The statistically validated improvements highlight BERT's potential in emotion-based video content analysis, demonstrating its superior ability to interpret emotional expressions in speech-derived text.

Table 6. F1 Score comparison.

Comparison	p-value
BERT vs. CNN	0.0006
BERT vs. LSTM	0.0010
BERT vs. SVM	0.0003
BERT vs LR	0.000

4.7 Discussion and Future work

The results of this study demonstrate that the pre-trained BERT model significantly outperforms traditional deep learning and machine learning approaches, achieving the highest accuracy, precision, recall, and F1 score across all evaluated models. One key aspect of model performance is the trade-off between precision and recall, particularly in the context of emotion analysis from speech-to-text data. While precision measures the proportion of correctly predicted positive instances among all predicted positives, recall evaluates the model's ability to correctly identify all relevant instances within the dataset. In highly imbalanced classification tasks, particularly in emotion recognition, an overemphasis on precision may lead to lower recall, meaning the model may fail to capture certain emotional expressions that do not appear frequently in the training data. Conversely, prioritizing recall may increase false positives, reducing the model's specificity in detecting nuanced emotions. Our results show that BERT maintains a strong balance between these metrics, minimizing the loss of relevant information while maintaining high classification

accuracy.

Despite the strong performance of the proposed framework, this study has certain limitations. First, the model relies on pre-trained embeddings [19], which may not fully capture the domain-specific nuances of emotion in social media videos. Fine-tuning BERT on a larger, domain-specific dataset could further enhance its contextual understanding of emotion intensities. Second, the computational complexity of BERT poses challenges for real-time inference, particularly in resource-constrained environments. Optimizations such as quantization, pruning, or knowledge distillation could be explored to make the model more efficient while preserving its classification performance. Additionally, our approach does not incorporate multimodal features such as facial expressions, tone of voice, or visual cues, which play a crucial role in emotional perception. Future work should integrate multimodal learning techniques by combining audio, visual, and textual data to enhance the robustness of emotion classification.

Another potential direction for improvement is handling ambiguous emotions, where a sentence may express multiple overlapping emotions. Future research could explore hierarchical classification techniques or attention-based mechanisms to dynamically capture the intensity and variations of emotions within social media content [20]. Moreover, the application of large language models (LLMs) such as GPT-based architectures or hybrid transformer models could be investigated to further refine emotion recognition accuracy. Lastly, conducting a more comprehensive statistical significance analysis using additional benchmarks and expanding the dataset to include diverse linguistic and cultural variations could further validate the model's generalizability and effectiveness in real-world applications.

5 Conclusion

With the rapid popularization of the Internet and the rapid development of multimedia processing technology, video data from different fields is increasing at an alarming rate. This tremendous expansion has greatly changed the world while the COVID-19 pandemic also contributes appreciably because of online classes, meetings, and conference recordings. Solving the challenges faced by this tremendous expansion of video with emotions is a burdensome task for the user. This research work proposed an interactive framework based on the Pre-Trained BERT model to classify the emotions

of video. The pre-trained BERT model is selected because of its higher results achieved by model evaluation criteria and performs well in the task of predicting emotions in comparison with other models. The Pre-Trained Model generates the output file that holds the information of emotions and applies a "maximum algorithm" on it to extract the most dominant emotions and maps it to social media selected videos for perspectives of viewers and boosts their number because of their phraseology to make him/herself more relaxed, which can significantly improve patient-provider communication and decision-making in clinical settings. Future research work can start from the text representation method to further improve the model's cross-domain ability to predict emotional intensity.

Data Availability Statement

Data will be made available on request.

Funding

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Pokhrel, S., & Chhetri, R. (2021). A Literature Review on Impact of COVID-19 Pandemic on Teaching and Learning. *Higher Education for the Future, 8*(1), 133-141. [CrossRef]

[2] Wilson, T., Wiebe, J., & Hoffmann, P. (2009). Articles: Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis. *Computational Linguistics, 35*, 399-433. [CrossRef]

[3] Palacios-Ceña, D., Fernández-de-Las-Peñas, C., Florencio, L. L., de-la-Llave-Rincón, A. I., & Palacios-Ceña, M. (2021). Emotional experience and feelings during first COVID-19 outbreak perceived by physical therapists: A qualitative study in Madrid, Spain. *International Journal of Environmental Research and Public Health, 18*(1), 127. [CrossRef]

[4] Silvia, P. J. (2009). Looking past pleasure: Anger, confusion, disgust, pride, surprise, and other unusual aesthetic emotions. *Psychology of Aesthetics, Creativity, and the Arts, 3*(1), 48-51. [CrossRef]

[5] Han, P., Chen, H., Rasool, A., Jiang, Q., & Yang, M. (2025). MFB: A Generalized Multimodal

Fusion Approach for Bitcoin Price Prediction Using Time-Lagged Sentiment and Indicator Features. *Expert Systems with Applications, 261*, 125515. [CrossRef]

[6] Lee, J., Jatowt, A., & Kim, K.-S. (2021). Discovering underlying sensations of human emotions based on social media. *Journal of the Association for Information Science and Technology, 72*(4), 417-432. [CrossRef]

[7] Kumar, A., Cambria, E., & Trueman, T. E. (2021, December). Transformer-based bidirectional encoder representations for emotion detection from text. In *2021 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1-6). IEEE. [CrossRef]

[8] Barken, R. (2013). Review of Emotions Matter: A Relational Approach to Emotions. *Symbolic Interaction, 36*(1), 104-106. [CrossRef]

[9] Rasool, A., Jiang, Q., Qu, Q., Kamyab, M., & Huang, M. (2022). HSMC: Hybrid Sentiment Method for Correlation to Analyze COVID-19 Tweets (pp. 991-999). [CrossRef]

[10] Landmann, H. (2021). The Bright and Dark Side of Eudaimonic Emotions: A Conceptual Framework. *Media and Communication, 9*(2), 191-201. [CrossRef]

[11] Illendula, A., & Sheth, A. (2019). Multimodal Emotion Classification. In *Companion Proceedings of The 2019 World Wide Web Conference* (pp. 439-449). ACM. [CrossRef]

[12] Saud, M., Mashud, M., & Ida, R. (2020). Usage of social media during the pandemic: Seeking support and awareness about COVID-19 through social media platforms. *Journal of Public Affairs, 20*(4), e02417. [CrossRef]

[13] Rasool, A., Shahzad, M. I., Aslam, H., & Chan, V. (2024). Emotion-Aware Response Generation Using Affect-Enriched Embeddings with LLMs. *arXiv preprint arXiv:2410.01306*. [CrossRef]

[14] Xu, G., Li, W., & Liu, J. (2020). A social emotion classification approach using multi-model fusion. *Future Generation Computer Systems, 102*, 347-356. [CrossRef]

[15] Păvăloaia, V.-D., Teodor, E.-M., Fotache, D., & Danileţ, M. (2019). Opinion Mining on Social Media Data: Sentiment Analysis of User Preferences. *Sustainability, 11*(16), 4459. [CrossRef]

[16] Han, P., Hong, J., Rasool, A., Chen, H., Pan, Y., & Jiang, Q. (2022, December). A hybrid recommendation model for social network services using twitter data. In *International Conference on Web Services* (pp. 122-129). Cham: Springer Nature Switzerland. [CrossRef]

[17] Ziyada, M., & Shamoi, P. (2024, November). Video Popularity in Social Media: Impact of Emotions, Raw Features and Viewer Comments. In *2024 Joint 13th International Conference on Soft Computing and Intelligent Systems and 25th International Symposium on Advanced Intelligent Systems (SCIS&ISIS)* (pp. 1-7). IEEE. [CrossRef]

[18] Heule, R., Bause, J., Pusterla, O., & Scheffler, K. (2020). Multi-parametric artificial neural network fitting of phase-cycled balanced steady-state free precession data. *Magnetic Resonance in Medicine, 84*(6), 2981-2993. [CrossRef]

[19] Min, B., Ross, H., Sulem, E., Veyseh, A. P. B., Nguyen, T. H., Sainz, O., ... & Roth, D. (2023). Recent advances in natural language processing via large pre-trained language models: A survey. *ACM Computing Surveys, 56*(2), 1-40. [CrossRef]

[20] Feng, B., Cheng, F., Liu, Y., Chang, X., Wang, X., & Jin, D. (2024). Community Detection on Social Networks With Sentimental Interaction. *International Journal of Semantic Web and Information Systems, 20*(1), 1-23. [CrossRef]

Muhammad Usman Ghani Received the BS(IT) degree from Superior University, Lahore, Pakistan. He is currently pursuing MS(CS) degree from National College of Business Administration & Economics Lahore, Pakistan. His research interests include deep learning, Natural language processing and LLMs. (Email: ug25192@gmail.com)

Umair Aziz Received the M(CS) degree from Air University, Islamabad, Pakistan. He is currently pursuing MS(CS) degree from National College of Business Administration & Economics Lahore, Pakistan. His research interests include deep learning, Natural language processing and LLMs. (Email: umairaziz3000@gmail.com)

Zubair Akbar (Student Member, IEEE) received the MS(CS) degree from Air University, Islamabad, Pakistan. His research interests include deep learning and human sleep monitoring. (Email: zobikakr96@gmail.com)