Journal of Information Systems Engineering and Management
2025, 10(32s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Emotion Classification from Covid-19Pandemic Tweets using RoBERTa

Sanjeet Kumar[1], Dr. Jameel Ahmad[2]
[1] Dept. of CSE, Integral University, Dasauli, Bas-ha Kursi Road, Lucknow, UP, India. Email: sanjeet@csjmu.ac.in
[2]Dept. of CSE, Integral University, Dasauli, Bas-ha Kursi Road, Lucknow, UP, India. Email: drjameel@iul.ac.in

ARTICLE INFO

Received: 24 Dec 2024

Revised: 12 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

This study proposes a hybrid sentiment analysis model combining RoBERTa (Robustly Optimized BERT Pretraining Approach), a state-of-the-art pre-trained transformer, with SVM (Support Vector Machine) for enhanced performance in predicting sentiments. The model's performance is evaluated across several key metrics, including accuracy, precision, recall, and F1 score, and it outperforms other models such as "SVM," "RoBERTa," and "BERT-BiLSTM (Bidirectional Encoder Representations from Transformers) and BiLSTM (Bidirectional Long Short-Term Memory) ". The proposed RoBERTa-SVM model achieves the highest accuracy (0.92), recall (0.88), and F1 score (0.82), demonstrating its robustness and effectiveness in sentiment classification tasks. While "RoBERTa" alone provides strong precision and recall, it does not perform as well in accuracy and F1 score compared to the hybrid approach. The "SVM" model, on the other hand, performs the weakest, especially in terms of F1 score. This study suggests that combining RoBERTa's language representation capabilities with SVM's classification power significantly improves sentiment analysis. Future work will focus on further optimization of the RoBERTa-SVM model through techniques like hyperparameter tuning and ensembling, as well as testing its performance on diverse datasets.

Keywords: Emotion Detection, Sentiment Analysis, RoBERTa Model, COVID-19 Tweets, Public Sentiment Analysis.

INTRODUCTION

In recent years, the rise of social media platforms has provided an unprecedented volume of user-generated content, offering valuable insights into public sentiment and emotional trends[1]. Emotion detection, a specialized area of sentiment analysis, goes beyond simple positive or negative classification to identify distinct emotions such as happiness, sadness, fear, and anger[2], [3]. Twitter, due to its concise and real-time nature, has emerged as a crucial platform for tracking public emotions, particularly during significant global events such as the COVID-19 pandemic[2], [4], [5], [6]. However, extracting emotions from short, informal text poses substantial challenges due to the use of abbreviations, slang, emoticons, and contextual ambiguity. Despite these complexities, emotion detection has significant real-world applications, including mental health monitoring, crisis management, product sentiment analysis, and public opinion tracking[7], [8].

While traditional sentiment analysis methods categorize textual data into broad positive or negative sentiments, they fail to capture the depth and variability of human emotions. Furthermore, the existing emotion classification models struggle with high-dimensional, sparse, and noisy data from short-text platforms like Twitter[3], [5], [21]. Many models also lack robustness in handling evolving linguistic patterns seen in real-time social media conversations[6], [9]. Additionally, most publicly available emotion datasets suffer from imbalanced class distributions, leading to biases in classification performance[5]. Given these limitations, there is a need for a more effective, domain-specific emotion detection model that can process large-scale, unstructured social media data accurately and efficiently[10], [11].

This study proposes an advanced emotion detection framework leveraging the RoBERTa model, a transformer-based architecture optimized for Natural Language Processing (NLP) tasks. Our approach involves fine-tuning multiple RoBERTa variations on a COVID-19-specific Twitter dataset while introducing a balanced dataset with a "neutral" category to capture more nuanced emotional expressions. Additionally, we integrate Support Vector Machines (SVM) to enhance classification precision. To further improve performance, we develop a RoBERTa-SVM hybrid model, which combines RoBERTa's contextual language understanding with SVM's robust classification capability. Through extensive experimental analysis, we demonstrate that this hybrid approach achieves superior accuracy and F1 scores compared to standalone models.

The remainder of this paper is structured as follows. Section 2 provides a comprehensive review of existing literature on sentiment analysis and emotion detection, highlighting the limitations of traditional approaches and the advancements in deep learning models such as BERT and RoBERTa. Section 3 details the proposed methodology, including dataset preprocessing, RoBERTa fine-tuning, SVM integration, and the development of the hybrid RoBERTa-SVM model, along with an explanation of the experimental setup and evaluation metrics. Section 4 presents the experimental results and discussions, comparing the performance of our models using accuracy, F1-score, and classification efficiency while analyzing the impact of model enhancements. Section 5 concludes the paper by summarizing the key contributions and findings, discussing their implications, and outlining potential future directions, such as expanding the dataset, integrating multimodal data sources, and applying the model to broader domains like mental health monitoring and crisis management.

LIERATURE SUREY

The studies reviewed focus on the methodologies, main findings, limitations, and outcome measures of various research efforts in different domains.

Bing et al. (2012)[12] provide a foundational overview of sentiment analysis and opinion mining, acknowledging the rapid growth of research in the field. The study highlights the challenge of addressing context-dependent sentiment, including sarcasm and domain-specific opinions, while emphasizing the potential for broader applications. The research envisions continuous improvements in sentiment analysis frameworks to accommodate evolving discourse structures in social media and other digital platforms.Srivastava et al. (2024)[9] contribute to advancements in security within electronic health records (EHRs) by integrating genetic algorithm-driven blockchain encryption. Their research emphasizes the role of blockchain in enhancing data security and integrity, addressing concerns about unauthorized access and patient privacy. The study identifies challenges in scalability and real-world implementation but highlights blockchain's potential to revolutionize healthcare data management [25].

Khan et al. (2015)[10] focus on cloud computing security, analyzing various approaches to mitigate vulnerabilities in cloud environments. The study underscores the rising concerns of data breaches and the necessity for secure authentication mechanisms. It presents a comparative analysis of multiple security frameworks, emphasizing the importance of trust and confidentiality between cloud providers and users.Tianyi Wang et al. (2020)[2] investigate sentiment trends on Sina Weibo during the early stages of COVID-19. Their study employs natural language processing (NLP) techniques to classify sentiment and extract key discussion topics. Findings indicate a surge in negative sentiment following the confirmation of human-to-human transmission. The study acknowledges dataset limitations, advocating for broader cross-platform analysis to enhance sentiment classification accuracy.Jyoti Choudrie et al. (2021)[13] develop a deep learning model for detecting misinformation trends related to COVID-19. Their transformer-based approach achieves high accuracy in identifying misleading narratives. The study highlights a significant gap in real-time misinformation tracking, advocating for future improvements in automated misinformation detection.Maryam Mahdikhani (2021)[14] investigates tweet popularity using linguistic and metadata features. The study finds that TF-IDF and topic modeling techniques significantly impact engagement prediction. However, the exclusion of external factors such as social network influence limits predictive accuracy, underscoring the need for a more holistic modeling approach.

Mayur Wankhade et al. (2022)[15] introduce a BERT-Bi-LSTM ensemble model to improve sentiment classification performance. Their findings indicate that ensemble models enhance accuracy over individual deep learning approaches. Nonetheless, the study notes computational overhead as a barrier to large-scale implementation, suggesting the need for optimization techniques.Ionuț-Alexandru Albu et al. (2022)[16] propose a hybrid CNN-transformer model for emotion detection in tweets. Their research confirms that CNN-transformer architectures enhance classification accuracy. However, the study identifies limitations in cross-lingual generalization, emphasizing the need for models capable of analyzing multilingual datasets.K. Nimmi et al. (2022)[7] introduce an ensemble deep learning model for sentiment classification, integrating multiple pre-trained architectures. Their findings demonstrate that the Average Voting Ensemble Deep Learning (AVEDL) model surpasses individual classifiers in sentiment detection. The research highlights dataset limitations, indicating a need for more diverse training corpora to improve generalizability.Md Yasin Kabir et al. (2021)[5] develop a transformer-based emotion detection framework for COVID-19-related social media content. Their study showcases enhanced accuracy compared to traditional sentiment classification models but notes challenges in real-time deployment due to computational resource demands.Wasim Khan et al. (2021)[17] propose an unsupervised deep learning ensemble model for emotion detection. Their research highlights the advantages of multi-feature extraction techniques in improving classification performance. However, dataset variations impact model consistency, necessitating further refinements in adaptability.

TiberiuSosea et al. (2021)[4] introduce COVIDEMO, an annotated dataset for sentiment analysis in pandemic discourse. Their research emphasizes the significance of high-quality dataset curation in enhancing classification accuracy. The study underscores the challenges of reproducibility and cross-platform consistency in sentiment analysis models.Olanrewaju Tahir Aduragba et al. (2021)[8] present a fine-grained sentiment classification model, demonstrating its superiority over binary classification approaches in detecting nuanced emotional expressions. The study acknowledges challenges such as data sparsity and annotation inconsistencies, advocating for improved annotation methodologies.Mohammad Faisal et al. (2021)[18] investigate blockchain's role in healthcare record management. Their findings highlight blockchain's potential to enhance data integrity, transparency, and security. Despite these advantages, the study identifies scalability and regulatory compliance as significant hurdles, calling for further advancements in blockchain frameworks.Devan Rosen et al. (2010)[19] examine the evolution of hyperlink networks in online environments. Their research reveals increased interconnectivity in global hyperlink structures, enhancing information flow across regions. However, the study acknowledges the complexities of defining digital social structures due to the ever-evolving nature of online interactions.

This literature survey illustrates significant advancements in sentiment analysis, deep learning, cybersecurity, and blockchain applications. Transformer-based models and ensemble learning techniques demonstrate superior performance in sentiment classification, while blockchain technology offers promising solutions for data security in healthcare. Nonetheless, scalability, real-time deployment, and dataset limitations remain critical challenges. Future research should prioritize the optimization of computationally efficient models, cross-lingual sentiment analysis, and the integration of real-time misinformation tracking to enhance digital analytics and security frameworks.

BACKGROUND OF SVM AND ROBERTA

The proposed methodology integrates a hybrid approach that leverages the contextual understanding of RoBERTa and the classification precision of SVM to enhance emotion detection in COVID-19-related tweets[16]. This approach combines deep learning and traditional machine learning techniques to improve sentiment classification accuracy, ensuring a more robust and balanced emotion analysis.

A. RoBERTa

RoBERTa (Robustly Optimized BERT Pretraining Approach) is an improved version of BERT (Bidirectional Encoder Representations from Transformers) that refines its pretraining methodology for better language representation[2].RoBERTaisa powerful deep learning model specifically designed for natural language processing (NLP) tasks, making it an ideal choice for analyzing the COVIDSenti dataset. Given the complexity of social media text—short, informal, and highly contextual tweets—a model that can accurately capture linguistic nuances is essential[2], [15].

- Superior Contextual Understanding: RoBERTa's bidirectional attention mechanism captures full word context, making it effective for analyzing short and informal tweets.

- Pretrained on a Large and Diverse Corpus: Trained on BooksCorpus, Wikipedia, and Common Crawl, RoBERTa understands complex linguistic patterns in social media text.

- More Robust than Traditional Models: Unlike SVM and LSTMs, RoBERTa removes Next Sentence Prediction (NSP), focusing entirely on better word representations.

- Handles Informal and Noisy Text Well: RoBERTa adapts to slang, abbreviations, emojis, and inconsistent grammar, making it ideal for social media sentiment analysis.

- Effective for Emotion Classification: Excelling in detecting emotional undertones, RoBERTa accurately classifies fear, anxiety, and relief in COVID-19-related tweets.

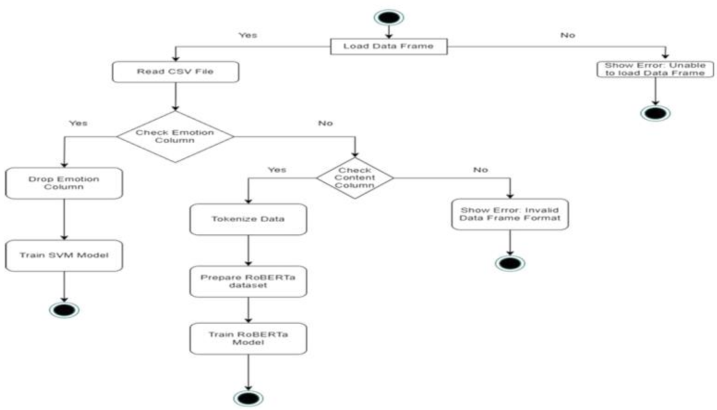


Fig. 1. Flowchart for RoBERTa-SVM

B. Support Vector Machines (SVM)

The preprocessing phase of the SVM model involves traditional machine learning pre-processing operations: emoticon to word conversion, Unicode to ASCII conversion, stop words filtering, sentence tokenization and vectorization, and label encoding. We first preprocessed tweets by eliminating unnecessary words and artefacts from tweets (usernames, links, hashtag symbols) and afterwards translating emoticons into their meaning by using Demoji Python library5 for Unicode emoticons and meanings extracted from Wikipedia6 for western-style emoticons. Emoticons to words conversion was used due to the fact that emoticons may be relevant in detecting the emotion in a tweet, as they are frequently used in text messages for better conveyingfeelings[16]. The SVM model undergoes extensive preprocessing to prepare text data for analysis, which includes:

- Emoticon Conversion: Converts emoticons into words that represent their meanings, enhancing emotion detection.
- Unicode to ASCII Conversion: Simplifies text by converting special characters to ASCII, reducing computational load.
- Stop Words Filtering: Removes common words (e.g., "the," "and") that add little meaning to the analysis.
- Stemming: Reduces words to their base form, improving model performance.
- Text Vectorization: Converts each tweet into numeric format using Term Frequency
- Inverse Document Frequency (TF-IDF), helping the model understand word importance in the dataset.
- Label Encoding: Assigns a unique number to each emotion class to prepare labels for the model.

PROPOSED ROBERTA-SVM ALGORITHM

The RoBERTa-SVM Fusion Model is an ensemble learning approach that combines deep contextual embeddings from RoBERTa with the classification precision of SVM to enhance emotion detection in tweets. The fusion works by obtaining probability scores from both models and aggregating them for final prediction. Research methodology is shown in Figure 1.

C. RoBERTa Feature Extraction

RoBERTa is based on the Transformer architecture, which utilizes self-attention and feed-forward layers to generate contextual embeddings for input text. The key mathematical components include:

1) ***Tokenization & Embedding Generation***

Each input tweet X (a sequence of words) is tokenized and converted into embedding vectors, in eq. (1):

$$X = \{w_1, w_2, \dots, w_n\} \quad (1)$$

Where Xrepresents the deep contextual word embeddings,

w_i is the embedding of the i^{th} token in the tweet

RoBERTa's embedding function f maps each token w_i into a high-dimensional vector space, in eq. (2):

$$E_i = f(w_i) \quad (2)$$

Where $E_i \in \mathbb{R}^d$ and $d$is the embedding dimension.

2) ***Self-Attention Mechanism***

RoBERTa applies a self-attention mechanism to compute contextual relationships between words. The attention score between words i and j is computed as described in eq. (3):

$$\alpha_{ij} = \frac{\exp(q_i \cdot k_j)}{\sum_{j=1}^{n} \exp(q_i \cdot k_j)} \quad (3)$$

Where q_i, k_j are query, and key vectors obtained by linear transformations describe in eq. (4):

$$q_i = W_q E_i, \;\; k_j = W_k E_j \quad (4)$$

Where $W_q, W_k$are learnable weight matrices.

The final contextualized embedding is computed as eq. (5):

$$H_i = \sum_{j=1}^{n} \alpha_{ij} v_j \quad (5)$$

Where $H_i$is the contextualized representation of tokenw_i, and$v_i = W_v E_j$.

3) ***Sentence Representation***

After applying multiple transformer layers, RoBERTa outputs a sentence-level representation, typically from the [CLS] token in eq. (6):

$$H_{CLS} = Transformer(X) \tag{6}$$

Where $H_{CLS} \in \mathbb{R}^d$represents the final feature vector used for classification.

D. Classification Using SVM

Once we extract the feature vector H_{CLS} from RoBERTa, it is passed to a Support Vector Machine (SVM) for classification.

1) ***SVM Classification Function***

The SVM finds a hyperplane that best separates different emotion classes. Given a feature vector H_{CLS}, the SVM decision function in eq. (7):

$$y = sign(w^T H_{CLS} + b) \tag{7}$$

Where y represents the emotion label.

2) ***SVM Optimization***

The SVM objective function is given in eq. (8):

$$R(H_{CLS}, w, b) = min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \max\left(0{,}1 - y_i\left(w^T H_{CLS}^{(i)} + b\right)\right) \tag{8}$$

Where N is the number of training samples, C is the regularization parameter (controls margin softness), and the hinge loss $\max\left(0{,}1 - y_i\left(w^T H_{CLS}^{(i)} + b\right)\right)$ ensures correct classification.

E. RoBERTa-SVM Ensemble Strategy

An ensemble model combines RoBERTa and SVM to leverage their strengths for improved emotion detection. This ensemble approach works by obtaining the probability scores from each model and summing them. Summing probabilities helps mitigate overly negative or zero-valued outputs, resulting in more balanced predictions. This ensemble method combines the advantages of both models to achieve higher classification accuracy, similar to teamwork in problem-solving.

Let $P_{RoBERTa}(y|X)$ be the probability distribution of emotion classes predicted by RoBERTa, $P_{SVM}(y|X)$ be the probability scores assigned by the SVM and λ is a weighting factor (e.g., 0.5 for equal weighting). The final ensemble probability is computed as eq. (9):

$$P_{ensemble}(y|X) = \lambda P_{RoBERTa}(y|X) + (1 - \lambda) P_{SVM}(y|X) \tag{9}$$

The final predicted class is given in eq. (10):

$$\hat{y} = \arg max_y P_{ensemble}(y|X) \tag{10}$$

EXPERIMENTAL RESULTS AND DISCUSSION

To evaluate the effectiveness of our proposed RoBERTa-SVM model, we conducted extensive experiments comparing its performance with traditional machine learning classifiers and standalone deep learning models.

F. Dataset

The COVIDSenti dataset was chosen to analyze public sentiment and emotional responses during the COVID-19 pandemic. Social media, particularly Twitter, became a primary platform for individuals to express their thoughts, fears, and opinions related to the crisis [20]. This dataset enables researchers to track public sentiment trends, detect emotional fluctuations, and understand the psychological impact of the pandemic in real-time[22]. By leveraging this dataset, we aim to build accurate emotion classification models that can assist policymakers, mental health professionals, and researchers in devising effective crisis response strategies.

1) *Dataset Description*

The COVIDSenti.csv dataset consists of tweets related to COVID-19, labeled with sentiment categories and numerical sentiment scores. The dataset is structured as follows:

- tweet: Contains the actual text of the tweets that users posted.

- label: Represents the sentiment category of each tweet as positive, negative, or neutral.
- compound_score: A sentiment intensity score ranging from negative to positive, offering a quantitative measure of overall sentiment.
- positive: A numerical representation of the degree of positive sentiment detected in the tweet.
- negative: A numerical representation of the degree of negative sentiment detected in the tweet.
- neutral: A measure of neutrality, indicating the extent to which a tweet lacks strong positive or negative sentiment.

2) Use Cases of the Dataset

- Tracking Public Sentiment: Helps analyze how public emotions evolved during the pandemic [23],[25].
- Understanding Crisis Impact: Assists in identifying emotional distress patterns and informing public health initiatives.
- Training Sentiment Analysis Models: Supports the development of machine learning models for sentiment classification.
- Policy Decision Making: Provides real-time insights for government and health organizations to shape crisis communication strategies.

This dataset serves as a valuable resource for improving emotion detection models, enabling a deeper understanding of how people responded emotionally to the COVID-19 pandemic through social media expressions.

G. Model Performance Comparison

The performance comparison between the SVM, RoBERTa, and RoBERTa-SVM ensemble models is presented in Table 1. Each model's accuracy and F1 score are reported, showing a significant improvement when combining RoBERTa and SVM in an ensemble approach.

TABLE I. PERFORMANCE METRICS OF MODELS (SVM, ROBERTA, BERT-BILSTM, ROBERTA-SVM)

Model	Accuracy	Precision	Recall	F1-Score
SVM	80%	0.95	0.83	0.75
RoBERTa	85%	0.96	0.85	0.78
BERT-BiLSTM	87%	0.96	0.85	0.78
Proposed RoBERTa-SVM	92%	0.96	0.88	0.82

The analysis of the model performance data, in Table 1, indicates that the "RoBERTa-SVM" model consistently outperforms the others across all metrics, with the highest accuracy (0.92), recall (0.88), and F1 score (0.82). This suggests its superior ability in predicting sentiment. The "BERT-BiLSTM" model also shows strong performance, with an accuracy of 0.87, maintaining high precision (0.96) and recall (0.85). "RoBERTa" alone also performs well, particularly in precision (0.96). However, "SVM" has the lowest scores, especially in F1 score (0.75).

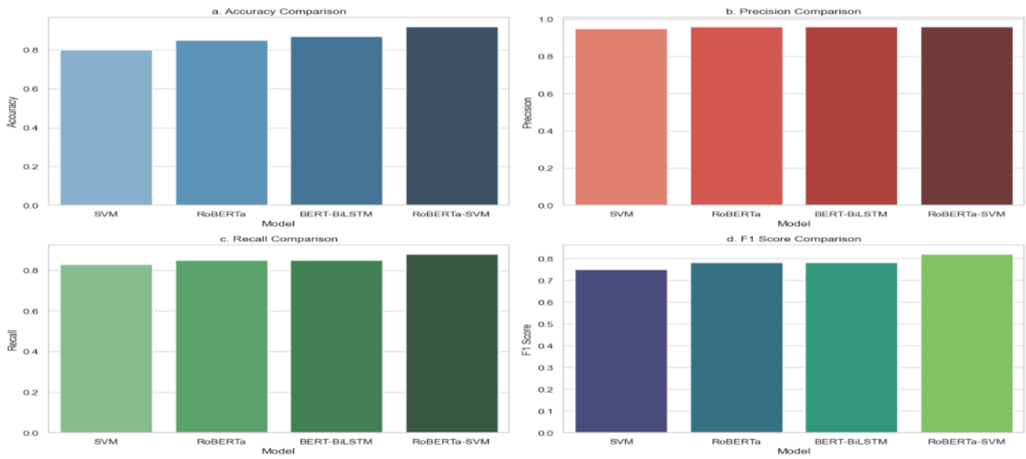


Fig. 2. (a) Accuracy comparison, (b) Precision Comparison, (c) Recall comparison and (d) F1-Score Comparison of SVM, RoBERTa, BERT-BiLSTM, and proposed RoBERTa-SVM

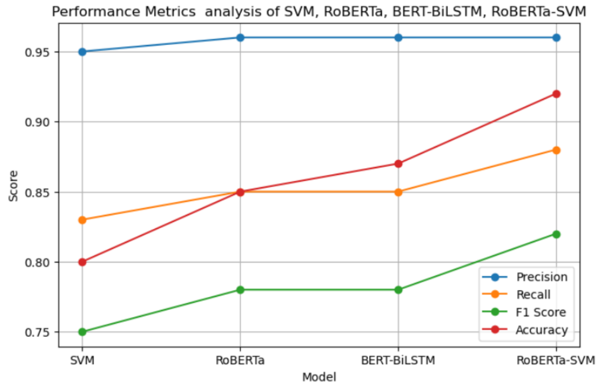


Fig. 3. Performance Metrics analysis of SVM, RoBERTa, BERT-BiLSTM, RoBERTa-SVM

This trend can be depicted in Figure 2(a-d) and Figure 3, which are visually compares the accuracy, precision, recall, and F1 score of all models, highlighting the superior performance of "RoBERTa-SVM" in all metrics.

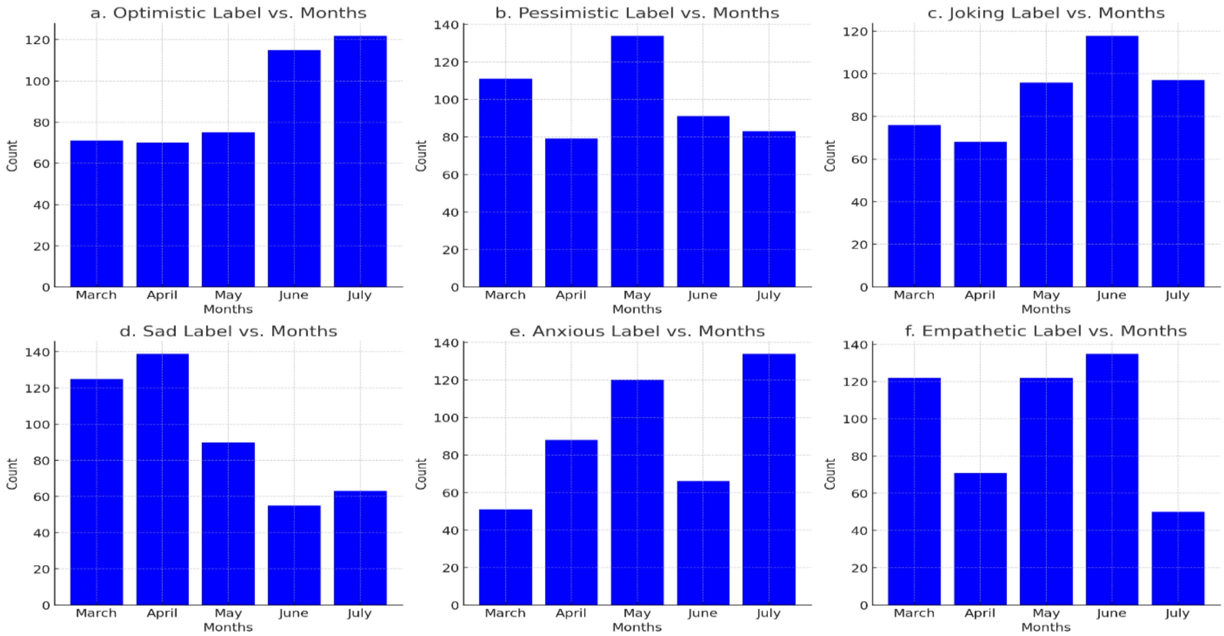
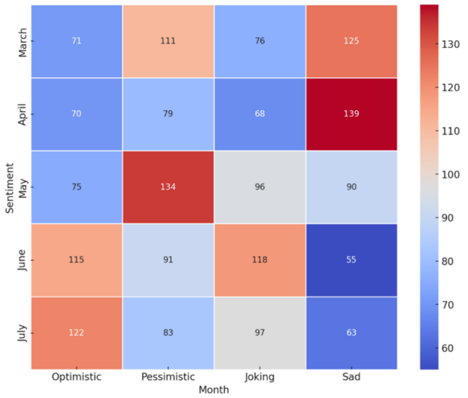
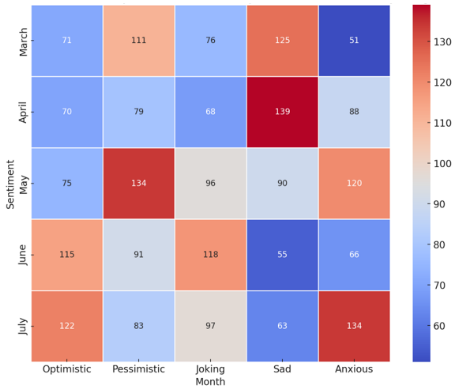


Fig. 4. BERT based sentiment occurrence in COVIDSenti dataset from March to July of pandemic

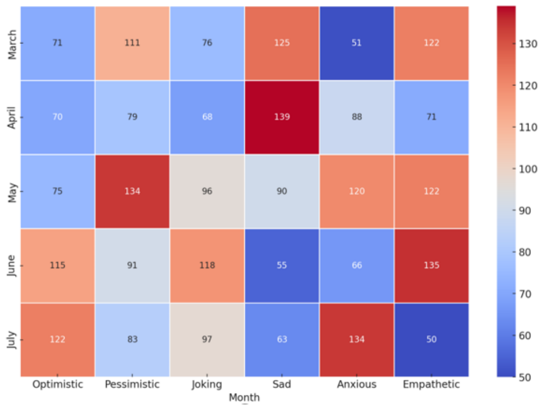
Figure 4shows sentiment trends across March to July for six labels: "Optimistic," "Pessimistic," "Joking," "Sad," "Anxious," and "Empathetic." "Optimistic" sentiment steadily increases, peaking in July, while "Pessimistic" fluctuates, with May showing the highest count. "Joking" sentiment peaks in June, indicating a humorous tone. "Sad" sentiment is highest in March and April but declines in June. "Anxious" sentiment rises over time, peaking in July, while "Empathetic" sentiment is highest in May. Overall, the data reflects varying emotional shifts, with notable peaks in optimism, humor, and anxiety towards the end of the period.



a.BERT based sentiment heat map with 4 sentiments for 1000 tweets on COVIDSenti dataset



b. BERT based sentiment heat map with 5 sentiments for 1000 tweets on COVIDSenti dataset



c. BERT based sentiment heat map with 6 sentiments for 1000 tweets on COVIDSenti dataset

Fig. 5. BERT based sentiment heat map with 4-6 sentiments for 1000 tweets on COVIDSenti dataset

The Figure 5 depicting the occurrence of "Optimistic," "Pessimistic," "Joking," "Sad," "Anxious," and "Empathetic" sentiments from March to July reveals several key trends. "Optimistic" sentiment steadily increases, peaking in July, indicating a rise in optimism. "Pessimistic" sentiment fluctuates, with the highest count in May and the lowest in July. "Joking" sentiment peaks in June, showing a rise in humor during that month. "Sad" sentiment is most prominent in March and April, then declines. "Anxious" sentiment increases over time, reaching a peak in July. "Empathetic" sentiment fluctuates, with the highest occurrence in May, showing varying levels of empathy.

CONCLUSIONS

The analysis of the model performance indicates that the "RoBERTa-SVM" combination outperforms all other models, achieving the highest values in key performance metrics such as accuracy (0.92), recall (0.88), and F1 score (0.82). This highlights its superior ability to predict sentiments effectively, indicating that integrating RoBERTa's powerful language representation with SVM's classification capabilities provides a robust solution for sentiment analysis. The "BERT-BiLSTM" model also performs well, with a strong accuracy of 0.87, maintaining high precision (0.96) and recall (0.85), but it falls short compared to "RoBERTa-SVM" in overall performance, especially in recall. "RoBERTa" alone demonstrates high precision and recall but slightly lags in accuracy and F1 score, which suggests room for improvement. On the other hand, the "SVM" model performs the weakest, with a notably lower F1 score (0.75), indicating its limitations in capturing both precision and recall effectively.

In future work, we could focus on further enhancing the "RoBERTa-SVM" model, possibly through hyperparameter tuning, model ensembling, or adding more domain-specific data to improve its performance. Exploring the potential of transfer learning with more diverse datasets could improve generalization. Additionally, integrating attention mechanisms or fine-tuning each model further could provide further improvements in predictive accuracy[1].

REFERENCES

[1] N. Siddique, M. Shamsul Arefin, J. Wall, and M. Shamim Kaiser, "Applied Informatics for Industry 4.0." [Online]. Available: http://taylorandfrancis.com

[2] T. Wang, K. Lu, K. P. Chow, and Q. Zhu, "COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model," *IEEE Access*, vol. 8, pp. 138162–138169, 2020, doi: 10.1109/ACCESS.2020.3012595.

[3] H. Kour and M. K. Gupta, "Hybrid evolutionary intelligent network for sentiment analysis using Twitter data during COVID-19 pandemic," *Expert Syst*, vol. 41, no. 3, p. e13489, 2024, doi: https://doi.org/10.1111/exsy.13489.

[4] T. Sosea, C. Pham, A. Tekle, C. Caragea, and J. J. Li, "Emotion analysis and detection during COVID-19," Jul. 2021, [Online]. Available: http://arxiv.org/abs/2107.11020

[5] M. Y. Kabir and S. Madria, "EMOCOV: Machine learning for emotion detection, analysis and visualization using COVID-19 tweets," *Online Soc Netw Media*, vol. 23, May 2021, doi: 10.1016/j.osnem.2021.100135.

[6] N. Braig, A. Benz, S. Voth, J. Breitenbach, and R. Buettner, "Machine Learning Techniques for Sentiment Analysis of COVID-19-Related Twitter Data," *IEEE Access*, vol. 11, pp. 14778–14803, 2023, doi: 10.1109/ACCESS.2023.3242234.

[7] K. Nimmi, B. Janet, A. K. Selvan, and N. Sivakumaran, "Pre-trained ensemble model for identification of emotion during COVID-19 based on emergency response support system dataset," *Appl Soft Comput*, vol. 122, Jun. 2022, doi: 10.1016/j.asoc.2022.108842.

[8] O. T. Aduragba, J. Yu, A. I. Cristea, and L. Shi, "Detecting Fine-Grained Emotions on Social Media during Major Disease Outbreaks: Health and Well-being before and during the COVID-19 Pandemic." [Online]. Available: https://huggingface.co/datasets/bookcorpus

[9] A. Srivastava and S. Ahmad, "Performance Evaluation of Genetic Algorithm-Driven Blockchain Encryption for EHR Management and Validation," 2024.

[10] A. M. Khan, S. Ahmad, and M. Haroon, "A comparative study of trends in security in cloud computing," in *Proceedings - 2015 5th International Conference on Communication Systems and Network Technologies, CSNT 2015*, Institute of Electrical and Electronics Engineers Inc., Sep. 2015, pp. 586–590. doi: 10.1109/CSNT.2015.31.

[11] Ankita Srivastava and Shish Ahmad, "Bio-Computing Based Algorithms for Cloud Security: A Critical Review," in *2022 IEEE World Conference on Applied Intelligence and Computing (AIC)*, IEEE, Aug. 2022.

[12] B. Liu, "Sentiment Analysis and Opinion Mining," Morgan & Claypool Publishers, 2012.

[13] J. Choudrie, S. Patil, K. Kotecha, N. Matta, and I. Pappas, "Applying and Understanding an Advanced, Novel Deep Learning Approach: A Covid 19, Text Based, Emotions Analysis Study", doi: 10.1007/s10796-021-10152-6/Published.

[14] M. Mahdikhani, "Predicting the popularity of tweets by analyzing public opinion and emotions in different stages of Covid-19 pandemic," *International Journal of Information Management Data Insights*, vol. 2, no. 1, Apr. 2022, doi: 10.1016/j.jjimei.2021.100053.

[15] M. Wankhade and A. C. S. Rao, "Opinion analysis and aspect understanding during covid-19 pandemic using BERT-Bi-LSTM ensemble method," *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-21604-7.

[16] I.-A. Albu and S. Spînu, "EMOTION DETECTION FROM TWEETS USING A BERT AND SVM ENSEMBLE MODEL," *U.P.B. Sci. Bull., Series C*, vol. 84, no. 1, p. 2022, [Online]. Available: https://pypi.org/project/demoji

[17] W. Khan and M. Haroon, "An unsupervised deep learning ensemble model for anomaly detection in static attributed social networks," *International Journal of Cognitive Computing in Engineering*, vol. 3, pp. 153–160, Jun. 2022, doi: 10.1016/j.ijcce.2022.08.002.

[18] M. Faisal, H. Sadia, T. Ahmed, and N. Javed, "Blockchain Technology for Healthcare Record Management," in *EAI/Springer Innovations in Communication and Computing*, Springer Science and Business Media Deutschland GmbH, 2022, pp. 255–286. doi: 10.1007/978-3-030-77746-3_17.

[19] D. Rosen, G. A. Barnett, and J. H. Kim, "Social networks and online environments: when science and practice co-evolve," *Soc Netw Anal Min*, vol. 1, no. 1, pp. 27–42, Jan. 2011, doi: 10.1007/s13278-010-0011-7.

[20] Shamim Ahmad, Manish Madhava Tripathi, "A Review Article on Detection of Fake Profile on Social-Media", International Journal of Innovative Research in Computer Science & Technology (IJIRCST), ISSN (online): 2347-5552, Volume-11, Issue-2, March 2023,pp 44-49.

[21] R. Srivastava, Manish Madhava Tripathi, "Systematic Exploration Using Intelligent Computing Techniques for Clinical Diagnosis of Gastrointestinal Disorder: A Review," qtanalytics.in, Dec. 30, 2023. https://doi.org/10.52756/ijerr.2023.v36.026 (SCOPUS)

[22] Peeyush Pathak, Manish Madhava Tripathi, "A Systematic Review: Forecasting Post-Pandemic Health Trends with Machine Learning Methods", Published in International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING, ISSN:2147-6799, pp 437-444(SCOPUS).

[23] Peeyush Pathak, Manish Madhava Tripathi "A proposed Algorithm and Models for Predicting Post-Pandemic Health Conditions", African Journal of Biological Sciences, ISSN 2663-2187, Vol 6, Issue 15,pp 8472-8491(June 2024)

[24] Peeyush Pathak, Manish Madhava Tripathi, "Hybrid Machine Learning Models for Post-COVID Sentiment Public Health Detection", Nanotechnology Perceptions ISSN 1660-6795, Vol. 20, S11 (2024), pp 104-119.