Literasi Nusantara
vol.6 no. 1, November 2025 – February 2026
https://journal.citradharma.org/index.php/literasinusantara/index

ISSN 2746 8208 (Online)
ISSN 2746 3575 (Print)

A Deep Learning Approach for Emotion Detection in Indonesian Social Media Texts

Rangga Widiasmara[1*], I Gede Susrama Mas Diyasa [2], Chrystia Aji Putra [3]

[1,2,3] Universitas Pembangunan Nasional Veteran, Jawa Timur, Surabaya, Indonesia

21081010085@student.upnjatim.ac.id , igsusrama.if@upnjatim.ac.id , ajiputra@upnjatim.ac.id

DOI : *https://doi.org/10.56480/jln.v6i1.1647*

Received: May 24, 2025
Revised: June 16, 2025
Accepted: July 20, 2025

Abstract

Automated emotion analysis of Indonesian social media text is challenging due to linguistic complexity and the nuanced use of emojis. This study addresses this by designing and evaluating a serial hybrid deep learning architecture combining a Convolutional Neural Network (CNN) and a Bidirectional Long Short-Term Memory (BiLSTM) network. The model was trained on an augmented Emotion Twitter dataset from the IndoNLU benchmark. Following a comprehensive preprocessing pipeline, the model achieved a promising test accuracy of 89.24% and a Macro F1-Score of 0.90 across five emotion classes. While the model excelled at identifying distinct emotions like 'love' (F1-Score of 0.96), it faced challenges in distinguishing between 'happy' and 'sadness'. These results establish the serial hybrid CNN-BiLSTM architecture as a viable and effective baseline for Indonesian emotion classification, providing a solid foundation for future research into more advanced models.

Keywords– *Emotion Classification, Deep Learning, CNN, BiLSTM, Indonesian Language*

1. Introduction

In the digital era, social media platforms like Twitter have become primary channels for expressing opinions and emotions. The ability to automatically analyze these emotions is invaluable for various applications, including brand monitoring, public health surveillance, and social science research (Singh et al., 2024). This task, known as emotion classification, is a subfield of sentiment analysis that aims to identify specific emotional states (e.g., happy, sad, anger) from text, moving beyond simple polarity (positive/negative) (Shaver et al., 2001).

The urgency for developing robust models for the Indonesian language is driven by its unique characteristics and the high volume of digital interaction. Indonesian text is marked by a rich morphology, a high prevalence of informal language, and slang, which complicates text normalization (Wilie et al., 2020). Furthermore, digital communication in Indonesia is often enriched with emojis, which serve as crucial non-verbal cues that can reinforce, alter, or even contradict the emotion conveyed by the text, adding another layer of complexity (Pane et al., 2022).

Deep learning models, particularly hybrid architectures that combine Convolutional Neural Networks (CNNs) and Bidirectional Long Short-Term Memory (BiLSTM), have demonstrated considerable promise in sentiment analysis research. CNNs are well-suited for extracting local features such as key phrases or patterns from textual data, while BiLSTMs excel at modeling long-range dependencies by processing sequences in both forward and backward directions (Phan et al., n.d.). The synergy between these two models allows for the simultaneous capture of spatial and temporal information, enabling a deeper understanding of textual semantics. This combination has been successfully implemented in various languages and domains, proving its robustness and versatility in emotion or sentiment classification tasks.

Despite the general support for this hybrid approach in international literature, a notable research gap remains—particularly regarding its application to Indonesian-language data. Most prior studies have gravitated toward developing increasingly complex models without first establishing a solid performance benchmark for basic CNN-BiLSTM architectures in this specific linguistic and cultural context. Moreover, few studies have accounted for the integration of non-textual cues such as emojis, which are prevalent in informal digital communication, especially on platforms like Twitter. Emojis often serve as emotional amplifiers or disambiguators and are thus crucial for accurately interpreting sentiment in social media discourse.

To address this gap, the present study adopts a focused approach to support and validate the performance of a serial CNN-BiLSTM hybrid model for emotion classification in Indonesian-language Twitter data. The core objective is to design, implement, and evaluate this architecture within a localized setting that includes both textual and emoji-based features. By doing so, the research aims to assess how well the model captures the nuances of Indonesian digital expressions, which are often informal, context-dependent, and rich in non-verbal indicators. This foundational analysis is essential for determining the feasibility of employing such hybrid models in real-world applications involving Indonesian-language sentiment data.

More broadly, this study contributes to the body of knowledge by establishing a reliable baseline that can inform the development of more advanced emotion classification models in the future. It offers a systematic evaluation framework and provides empirical evidence regarding the applicability of CNN-BiLSTM models in multimodal, culturally specific contexts. By anchoring its analysis in the Indonesian language and incorporating emoji cues, the study not only enhances the linguistic diversity of sentiment analysis research but also responds to the growing demand for culturally aware and context-sensitive artificial intelligence systems. The findings are expected to support the advancement of emotionally intelligent applications tailored to the Indonesian digital ecosystem.

2. Method

The dataset utilized in this study is sourced from the Emotion Twitter Dataset within the Indonesian Natural Language Understanding (IndoNLU) benchmark repository. This corpus comprises 4,403 Indonesian-language tweets, each manually annotated with one of five distinct emotional labels: anger, fear, happy, love, and sadness. To enhance the realism and representativeness of modern digital communication, the dataset was augmented by programmatically inserting between one to three emojis into each tweet. These emojis were selected based on their semantic relevance to the associated emotion label. The augmentation process was designed to simulate real-world Twitter discourse, where users frequently employ emojis as expressive tools that complement textual content and convey affective meaning that may be ambiguous in plain text.

Prior to model training, the dataset underwent a comprehensive preprocessing pipeline to ensure quality, consistency, and semantic richness. Initially, data cleaning steps were applied to remove duplicate entries and rows with missing or corrupted values. Text normalization followed, starting with case folding, in which all text was converted to lowercase to reduce dimensionality and avoid treating word variants differently. Non-essential characters such as URLs, user mentions, hashtags, and non-alphanumeric symbols were removed to eliminate noise. A key preprocessing step involved emoji conversion, wherein each emoji was translated into its corresponding Indonesian textual description, enabling the model to process their semantic content directly within the language modeling framework. Additional normalization was performed by standardizing slang and colloquial terms using a domain-specific dictionary—this step was particularly important given the informal and abbreviated nature of social media language. Finally, the cleaned texts were tokenized and padded to a uniform sequence length of 37 tokens, providing consistent input size across samples. The dataset was then split using a stratified strategy into training (80%), validation (10%), and test (10%) sets, ensuring balanced class representation across all partitions.

The neural architecture implemented in this study is a serial hybrid model combining CNN and BiLSTM layers, optimized for sequential emotion classification tasks. The model begins with an Embedding Layer that transforms tokenized word indices into 300-dimensional dense vectors using pre-trained FastText embeddings specific to the Indonesian language. Notably, the embedding weights were frozen during training to retain the semantic integrity of the pre-learned representations. This is followed by a 1D Convolutional Layer consisting of 128 filters with a kernel size of 3, which extracts local n-gram patterns from the embedded sequences, capturing meaningful word combinations that frequently correspond to emotional expressions. A MaxPooling1D layer then reduces the feature dimensionality, focusing on the most prominent local features. The condensed representation is subsequently fed into a Bidirectional LSTM layer with 256 units. This layer processes the temporal dependencies in both forward and backward directions, allowing the model to understand the contextual flow of language, which is crucial for capturing subtleties in emotional tone. The final output is passed through a Dense layer with a softmax activation function to generate probability distributions over the five emotion classes. To prevent overfitting and enhance generalizability, Dropout layers were applied at various stages in the architecture, including after pooling, BiLSTM, and dense layers.

The training phase was conducted using the Adam optimizer with an initial learning rate of 0.0005, and the categorical cross-entropy loss function was employed given the multi-class classification task. To address the class imbalance present in the dataset—particularly with emotions such as love and fear that occurred less frequently—a class weighting strategy was implemented. This technique assigns greater penalties to misclassifications of underrepresented classes, thereby encouraging the model to learn more balanced representations across categories. Training was performed using a batch size of 64, and to ensure training efficiency, two callbacks were used: EarlyStopping with a patience of 10 epochs to halt training once performance plateaued, and ReduceLROnPlateau to dynamically lower the learning rate in response to stagnating validation loss. Model performance was evaluated using four key metrics: accuracy, precision, recall, and F1-score. Crucially, a Macro-Average F1-score was used to ensure that each emotion class was treated equally, thereby providing a balanced and fair assessment of the model's generalization ability across all emotional categories.

3. Result and Discussion

The trained serial hybrid CNN-BiLSTM model was evaluated on the test set of 437 samples. The model achieved an overall test accuracy of 89.24% and a Macro F1-Score of 0.90. This indicates that the architecture is highly effective and provides a well-balanced performance for the five-class emotion classification task.

The detailed per-class performance is presented in Table 1.

Table 1. Classification Report for the Serial Hybrid CNN-BiLSTM Model

Emotion	Precision	Recall	F1-Score	Support
Anger	0.93	0.91	0.92	110
Fear	0.85	0.89	0.87	64
Happy	0.88	0.84	0.86	101
Love	0.98	0.94	0.96	63
Sadness	0.85	0.90	0.87	99
Accuracy			0.89	437
Macro Avg	0.90	0.90	0.90	437
Weighted Avg	0.89	0.89	0.89	437

The results show that the model performs exceptionally well on the love class, achieving an F1-Score of 0.96. This suggests that the linguistic and emoji cues for this emotion are very distinct and easily captured by the model. The anger class also shows strong performance (F1-Score of 0.92). The model's performance is lowest, though still robust, for the happy (0.86), fear (0.87), and sadness (0.87) classes.

To further analyze the model's behavior, a confusion matrix was generated (Figure 1).

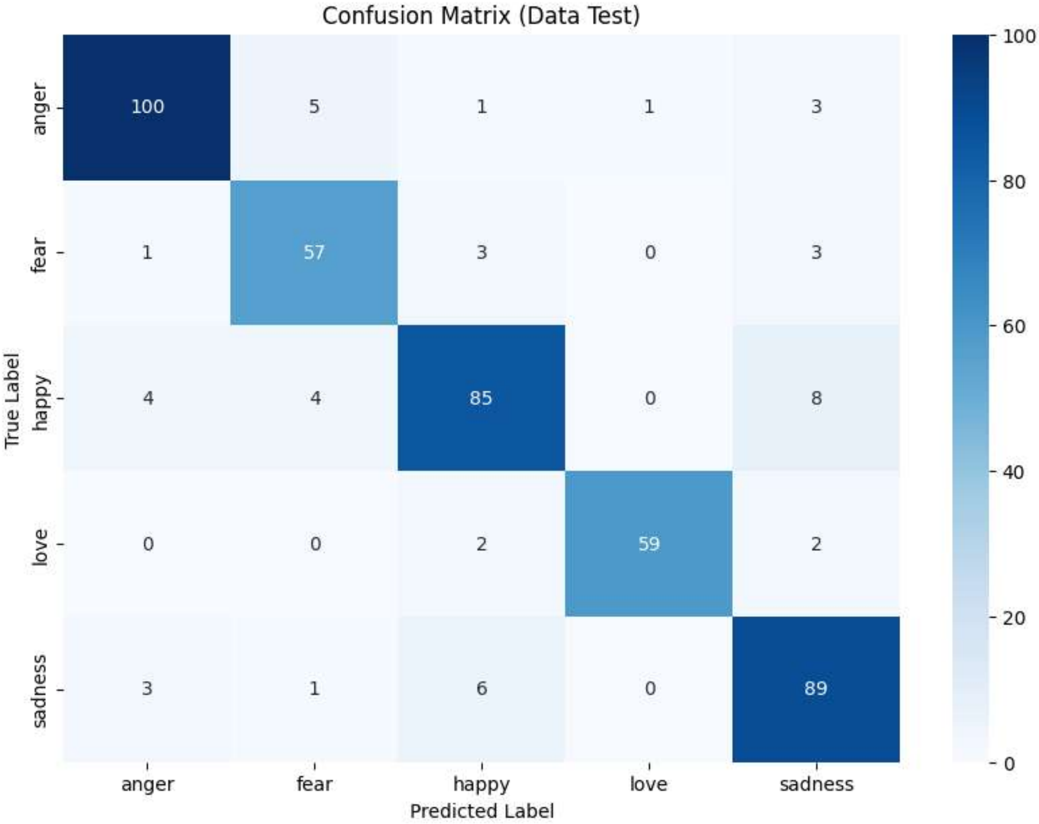

Figure 1. Confusion Matrix of Model Predictions on the Test Set

The confusion matrix obtained from the evaluation phase provides a detailed insight into the classification behavior of the serial hybrid CNN-BiLSTM model. The dominance of high values along the diagonal of the matrix confirms that the model achieved a strong overall classification performance, with most predictions aligning correctly with the true labels. For instance, the model accurately identified 100 out of 110 instances labeled as anger and correctly classified 59 out of 63 samples expressing love. This high level of accuracy across multiple categories suggests that the model effectively captured the distinguishing lexical and contextual patterns associated with those emotions, both from the text and the augmented emoji representations.

Despite the generally high accuracy, the confusion matrix also exposes specific areas where the model struggles to differentiate between emotionally adjacent classes. A notable source of error was observed in the misclassification between happy and sadness. Specifically, 8 instances labeled as happy were incorrectly classified as sadness, while 6 samples expressing sadness were misclassified as happy. This bidirectional confusion indicates a nuanced challenge in distinguishing these emotions, which often share overlapping vocabulary or arise in emotionally complex expressions. Such misclassifications underscore the limitations of relying solely on surface-level textual features, even when combined with emoji semantics, to disambiguate subtle emotional cues.

This pattern of confusion may be attributed to the intricacies of the Indonesian language and the informal, often figurative nature of social media discourse. In platforms like Twitter, users frequently express feelings using idiomatic phrases, sarcasm, or emotionally ambiguous statements. For example, expressions like "aku terharu" (I am moved) or "air mata bahagia" (tears of joy) can semantically straddle the boundary between happiness and sadness, depending on context. Without additional pragmatic or situational context, even a robust deep learning model may struggle to resolve such ambiguities. Furthermore, the emotional tone of a tweet may shift depending on the user's intent, cultural context, or the accompanying emoji, which may either reinforce or contradict the textual message.

Nevertheless, the overall results affirm that the serial hybrid CNN-BiLSTM model is capable of learning meaningful emotional representations from Indonesian-language Twitter data, particularly when textual input is enhanced with semantically integrated emoji features. The strong classification performance—combined with the model's ability to handle the informal and expressive nature of digital Indonesian—demonstrates its potential as a reliable benchmark for future research. By highlighting both its strengths and specific challenges, the confusion matrix analysis provides a roadmap for further refinement, such as integrating contextual embeddings or attention mechanisms to better capture subtle emotional nuances. Ultimately, this study establishes a foundational model for emotion classification in a low-resource language setting, contributing valuable insights to the broader field of sentiment analysis in multilingual and multimodal contexts.

4. Conclusion

This study successfully designed and evaluated a serial hybrid CNN-BiLSTM model for emotion classification on Indonesian Twitter data. The model achieved a high accuracy of 89.24% and a Macro F1-Score of 0.90, confirming that this architecture is a robust and effective method for tackling the complexities of the Indonesian language and integrated emoji use. While the model demonstrated excellent performance, particularly for distinct emotions like 'love', it also highlighted the persistent challenge of differentiating between semantically closer emotions like 'happy' and 'sadness'. This work provides a solid benchmark and a valuable foundation for future research, which could explore more advanced architectures, such as parallel or multi-input models, to further resolve these ambiguities.

References

Madan, A., & Kumar, D. (2024). Real-time topic-based sentiment analysis for movie tweets using hybrid approach. *Knowledge and Information Systems*, *66*(7), 3061–3083. https://doi.org/10.1007/s10115-024-02298-x

Pane, S. F., Ramdan, J., Putrada, A. G., Fauzan, M. N., Awangga, R. M., & Alamsyah, N. (2022). A hybrid CNN-LSTM model with word emoji embedding for improving the Twitter sentiment analysis on Indonesia’s PPKM policy. Dalam *2022 International Conference on Information Technology Systems and Innovation (ICITSI)*. IEEE. https://doi.org/10.1109/ICITSI56531.2022.10057720

Phan, H. T., Seo, Y.-S., & Nguyen, N. T. (t.t.). *Fuzzy hybrid CNN-LSTM model for sentence-level sentiment analysis*. SSRN. https://doi.org/10.2139/ssrn.4994871

Riyadi, S., Andriyani, A. D., & Sulaiman, S. N. (2024). Improving hate speech detection using double-layers hybrid CNN-RNN model on imbalanced dataset. *IEEE Access*, *12*, 159660–159668. https://doi.org/10.1109/ACCESS.2024.3487433

Shaver, P. R., Murdaya, U., & Fraley, R. C. (2001). Structure of the Indonesian emotion lexicon. *Asian Journal of Social Psychology*, *4*(3), 201–224. https://doi.org/10.1111/1467-839X.00086

Singh, A. K., Bhushan, A., & Dwivedi, D. (2024). Analyzing sentiments on Twitter using deep learning techniques. *International Journal of Modern Education and Computer Science*, *16*(1), 60–72. https://doi.org/10.5815/ijmecs.2024.01.05

Wilie, B., Vincent, T. N., Cahyawijaya, S., Li, X., Lim, Z. Y., Sutiono, A., ... & Purwarianti, A. (2020). IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding. Dalam *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.aacl-main.85