SIEmo-LSTM: Multimodal Fusion of Text, Unicode Emoji, and ASCII Emoticons for Indonesian Sentiment Analysis

M. Noer Fadli Hidayat
Department of Electrical Engineering and Informatics, Universitas Negeri Malang, Indonesia | Informatics Engineering Department, Faculty of Engineering, Universitas Nurul Jadid, Indonesia
m.noer.2305349@students.um.ac.id

Didik Dwi Prasetya
Department of Electrical Engineering and Informatics, Universitas Negeri Malang, Indonesia
didikdwi@um.ac.id (corresponding author)

Triyanna Widiyaningtyas
Department of Electrical Engineering and Informatics, Universitas Negeri Malang, Indonesia
triyannaw.ft@um.ac.id

Received: 7 January 2026 | Revised: 12 February 2026 and 25 February 2026 | Accepted: 27 February 2026

 DOI: https://doi.org/10.48084/etasr.17398

ABSTRACT

Sentiment analysis of Indonesian user reviews is challenged by informal language and frequent nonverbal cues such as Unicode emojis and ASCII emoticons. This study aimed to quantify the benefit of explicitly modeling ASCII emoticons together with Unicode emojis and text for three-class sentiment classification (negative/neutral/positive). SIEmo-LSTM is a tri-modal pipeline that (i) maps Unicode emojis using an emoji sentiment resource, (ii) detects, normalizes, and converts ASCII emoticons into descriptive tokens using Emot and LEED (93.3% successful conversion), and (iii) encodes the unified sequence using IndoBERT as a contextual feature extractor and refines it with a Bi-LSTM layer before multiclass prediction. Experiments used 304,570 Ruangguru app reviews (2016–2023), a tri-modal subset of 2,527 reviews, and a 70/20/10 train/validation/test split. Class imbalance was addressed using Random OverSampling (ROS). The full Text+SE+IE configuration with ROS achieved up to 0.9935 Accuracy and 0.9967 Macro-F1, outperforming text-only and text+Unicode baselines, while Random UnderSampling (RUS) consistently degraded performance. These findings imply that treating ASCII emoticons as a first-class affective modality—alongside Unicode emojis and text—improves robustness and class-balanced sentiment recognition for Indonesian user-generated reviews.

Keywords-sentiment analysis; multimodal fusion; unicode emoji; ASCII emoticons; IndoBERT; BiLSTM

I. INTRODUCTION

Sentiment analysis is widely used to extract user attitudes from large-scale online text, enabling organizations to monitor service quality and support decision-making. In Indonesia, sentiment mining has been applied to diverse domains (e.g., digital banking, political discourse, citizen-report classification, and policy-related debates) [1-4], and also to application-oriented review analytics such as hotel reviews [5]. A recent work on Indonesian TikTok comments further indicates that robust preprocessing for noisy and informal language can substantially improve performance, where FastText+BiLSTM with slang-to-formal translation achieved up to 93.10% test accuracy [6].

However, sentiment analysis remains challenging because reviews are often short, informal, and noisy (slang, abbreviations, creative spelling), making affective intent implicit rather than explicitly lexical; thus, sequential baselines such as LSTM-based classifiers remain relevant [7]. In parallel, transformer-based contextual encoders, such as BERT, have become foundational for NLP and are frequently integrated into sentiment pipelines due to strong downstream performance [8, 9]. BERT-based models can outperform various deep networks in domain-specific sentiment settings. Recent surveys also emphasize the evolution toward modern feature learning, integration approaches for robustness in short-text settings, embedding strategies, deep architectures, and evaluation metrics [10].

Figure 1 summarizes the multimodal pipeline adopted in this work, from preprocessing and cue normalization to unified sequence construction for model input.

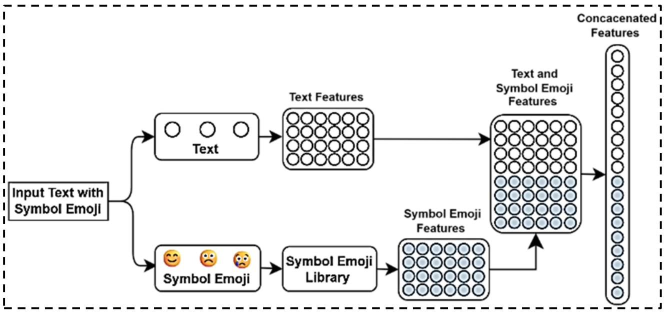

Fig. 1. Multimodal text and emoji pipeline.

A key limitation of text-only modeling is that users frequently express emotion through nonverbal markers—particularly emojis and emoticons. Emojis function as nonverbal communication cues [11], exhibit measurable sentiment tendencies [12], and can be treated more consistently using normative resources such as the Lisbon Emoji and Emoticon Database (LEED) for both emojis and emoticons [13]. Previous works integrated emoji cues into sentiment systems through fusion strategies and representation learning, reporting improvements over text-only settings [14-17]. Emoticons also strengthen short-text sentiment classification [18] and can improve neural pipelines when injected as explicit affective features [19]. A closely related study is emoji-text fusion via early/late fusion [14], while other works benchmarked text-emoji modeling [15], improved sentiment using emoji embeddings [16], and modeled emoji-text interactions with attention mechanisms (e.g., bi-sense emoji embedding with attention-LSTM) [20]. Emoticon-aware research shows that blending emoticons with short texts improves sentiment classification [18], including neural pipelines that inject emoticons as explicit features (e.g., attention-based CNN) [21], as well as classical alternatives such as naive Bayes with emoticon attributes [21] and dedicated text+emoticon sentiment formulations [22]. However, since most previous approaches remained bimodal (text+Unicode emoji or text+emoticon) and often normalized ASCII emoticons away, end-to-end tri-modal evaluation on Indonesian app reviews with controlled ablations is still limited.

To address this measurable gap, this paper proposes SIEmo-LSTM, a tri-modal sentiment framework that fuses Indonesian text, Unicode emojis, and ASCII emoticons for sentiment classification on Indonesian user reviews. This approach is motivated by evidence that emoji-text fusion improves opinion mining [14], emoticons contribute useful sentiment cues in short-text settings [18, 19], and transformer-based contextual encoders provide strong representations for sentiment modelling [8]. The main contributions of this study are:

- A tri-modal formulation for Indonesian sentiment analysis by jointly modelling text, Unicode emojis, and ASCII emoticons [14, 18]
- A practical multimodal pipeline that integrates nonverbal cues as consistent features for neural sentiment classification [13, 19]
- An empirical evaluation that demonstrates the value of incorporating emoticons alongside emojis and text for improved sentiment prediction [16, 17].

Table I provides a compact comparison of representative studies, highlighting how SIEmo-LSTM differs by jointly modeling Indonesian text, Unicode emojis, and ASCII emoticons and quantifying each modality's contribution under consistent ablation and imbalance-aware settings.

TABLE I. COMPARISON STUDY LITERATUR

Ref.	Task (Context)	Modalities	Main idea	Limitation vs. SIEmo-LSTM
[5]	Sentiment / Indonesian reviews	Text	Deep learning pipeline for Indonesian reviews	Nonverbal cues not central
[7]	Sentiment / Indonesian	Text	LSTM baseline for Indonesian sentiment classification	No emoji/emoticon modality
[13]	Resource	Emojis + emoticons	Norms for emojis/ emoticons (LEED)	Resource only; not an end-to-end classifier
[14]	Opinion mining	Text + Unicode emoji	Early/late fusion of emojis and text	Bimodal (no explicit ASCII emoticons)
[16]	Sentiment	Text + Unicode emoji	Emoji embedding improves sentiment accuracy	Bimodal; emoticons not modeled
[18]	Short-text sentiment	Text + emoticons	Blending emoticons with short texts	No Unicode emojis; not tri-modal
[19]	Microblog sentiment	Text + emoticons	Emoticons added to attention-based CNN	No Unicode emojis; not tri-modal
[20]	Twitter sentiment	Text + Unicode emoji	Bi-sense emoji embedding + attention-LSTM	Unicode-emoji centric; not tri-modal
[21]	Microblog multimedia sentiment	Text + emoticon attributes	Naive Bayes with emoticon attribute features	Classical model; limited contextual encoding
[23]	Sentiment method	Text + emoticons	Joint text emoticon sentiment formulation	No Unicode emojis; not tri-modal
This study	Indonesian app-review sentiment	Text + Unicode emoji + ASCII emoticons	Tri-modal fusion combines: (i) transformer-based text encoding [8], (ii) Unicode emoji sentiment cues, and (iii) normalized ASCII emoticon tokens, followed by BiLSTM sequence modeling and classification.	Addresses the tri-modal gap by jointly modeling Unicode emojis and ASCII emoticons

II. RESEARCH METHODOLOGY

Figure 2 illustrates the research framework for sentiment analysis.

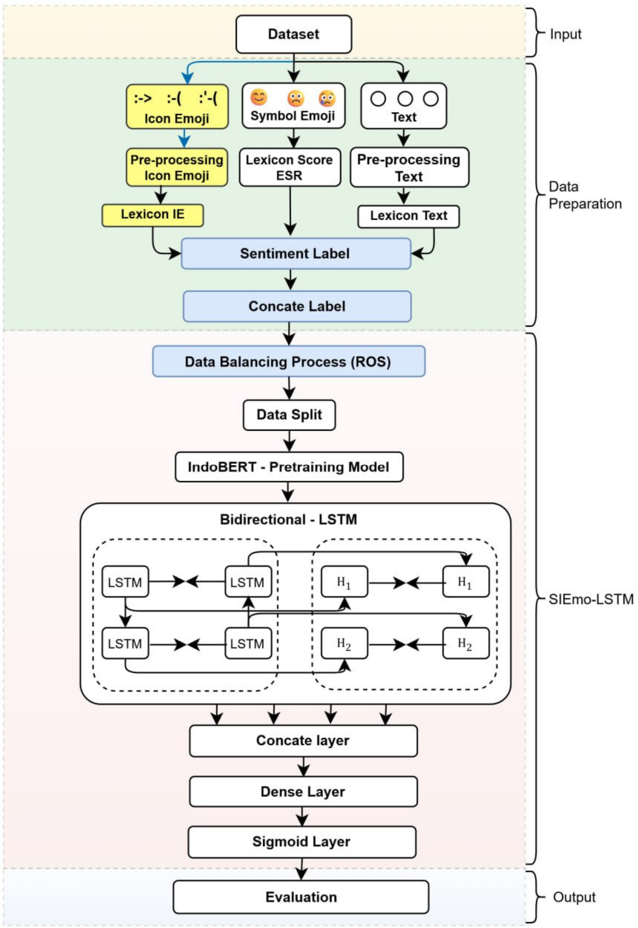

Fig. 2. SIEmo-LSTM end-to-end pipeline for Indonesian multimodal sentiment analysis (T+SE+IE).

A. *Dataset*

The raw corpus consists of 304,570 user reviews collected from a public review page in the Play Store of the Ruangguru mobile application (2016–2023). Reviews were acquired using an automated collection or crawling/scraping procedure on publicly accessible review pages, capturing the review text and available metadata. Data were collected for research purposes only. Rate-limiting was applied during collection, and any personally identifiable information was removed/ignored to ensure privacy.

A tri-modal subset was then built by retaining only reviews that contain Indonesian text alongside at least one Unicode emoji (SE) and at least one ASCII emoticon/icon emoji (IE) detected by the preprocessing pipeline, yielding 2,527 reviews for modeling. All reviews are treated as user-generated text. Any personally identifiable information (if present) was removed, and results are reported only in an aggregated form.

To improve data quality and replicability, filtering rules were applied before modeling: (i) remove empty/very short reviews; (ii) remove duplicate entries; (iii) normalize noisy patterns (repeated characters, spacing, and punctuation); and (iv) retain only Indonesian-language reviews using a consistent language check (or equivalent heuristic). After filtering, Unicode emojis were mapped using ESR, while ASCII emoticons were normalized and converted to descriptive tokens using emot and LEED, achieving 93.3% successful conversion. Reviews with ambiguous/unconvertible emoticons were conservatively assigned neutral handling to reduce bias.

B. *Overview of the Proposed Pipeline*

As illustrated in Figure 2, SIEmo-LSTM is designed as an end-to-end multimodal pipeline that jointly models Indonesian text (T), Unicode symbol emojis (SE), and ASCII icon emojis/emoticons (IE). The pipeline consists of four main stages. First, the input reviews are collected and filtered to retain multimodal instances containing the required modalities. Second, during data preparation, each modality is processed through a dedicated preprocessing path: the text is cleaned/normalized and labeled using an Indonesian sentiment dictionary [23], Unicode emojis are labeled using Emoji Sentiment Ranking (ESR) [12], and ASCII emoticons are normalized and converted into descriptive tokens using the emot library and LEED modeling [13], followed by sentiment labeling via Multilingual BERT (BERT-based) [8]. Third, the modality-wise sentiment signals are combined into a unified target through unweighted score-level late fusion (concate label), and class imbalance is mitigated using Random Oversampling (ROS) before learning. Finally, the SIEmo-LSTM model encodes the converted multimodal sequence using IndoBERT (BERT-based) as a contextual feature extractor [8], and then refines sequential dependencies with a BiLSTM layer (LSTM family) [24] before producing the final sentiment prediction.

C. *Multimodal Preprocessing and Modality-Specific Labeling*

For text, preprocessing follows standard NLP preparation for noisy user reviews, including cleaning and normalization so that non-standard patterns (e.g., informal tokens) are reduced before modeling. The sentiment signal of the text is obtained using the Indonesian sentiment dictionary [23], which provides the polarity basis for the text channel used in the fusion stage.

For Symbol Emojis (SE), Unicode emojis are labeled using Emoji Sentiment Ranking (ESR) v1.0 [12]. In this step, the emojis found in a review are mapped to ESR-based sentiment scores/labels. When an emoji is not available in the ESR mapping, the system assigns a conservative neutral label to avoid injecting ungrounded polarity.

For Icon Emojis/emoticons (IE), ASCII emoticons are first detected and normalized (e.g., removing repetitions and correcting common typo variants), then converted into descriptive word tokens using the emot library and LEED norms modeling [13]. After conversion, the sentiment label for the resulting emoticon-text representation is obtained using a Multilingual BERT sentiment classifier [8]. In cases where emoticons remain ambiguous or cannot be converted reliably, they are treated as neutral to minimize interpretation bias.

D. Concatenated Label via Unweighted Score-Level Late Fusion

To produce a single target label that reflects combined affective signals, SIEmo-LSTM uses unweighted score-level late fusion across modalities. Each modality label is mapped to a numeric score (negative = −1, neutral = 0, positive = +1), and for each review i the fused score is computed as:

$$S_i = S_i^{(T)} + S_i^{(SE)} + S_i^{(IE)} \quad (1)$$

The final label follows the sign of S_i: positive if $S_i > 0$, negative if $S_i < 0$, and neutral if $S_i = 0$. This unweighted formulation is adopted as a transparent and interpretable decision-level baseline that avoids arbitrary weighting when modality reliability varies in informal reviews while enabling controlled ablation across configurations (Text Only, Text+SE, Text+SE+IE, IE Only). The novelty lies not in the summation itself, but in explicitly treating ASCII emoticons as a first-class modality via normalization/conversion into consistent tokens, integrating Unicode emojis and emoticons in a unified pipeline, and evaluating modality contributions under multiple balancing strategies.

E. Data Balancing and Data Split

Since multimodal online-review datasets often exhibit class imbalance, SIEmo-LSTM applies ROS on the fused-label training data to reduce bias toward the majority class. This step duplicates minority-class samples until class sizes become more comparable, helping stabilize training and preventing "accuracy paradox" behavior where a model appears strong while failing minority classes. After concatenated labeling, the dataset is split into 70% training, 20% validation, and 10% testing. The validation set is used to monitor learning dynamics during training without updating weights, while the test set is reserved strictly for final evaluation to preserve an unbiased estimate of generalization performance.

F. SIEmo-LSTM Architecture: IndoBERT Feature Extraction + BiLSTM Sequence Modeling

SIEmo-LSTM represents each review as a unified token sequence where Indonesian text and the converted emoji/emoticon cues are inserted as semantic tokens, allowing all modalities to be processed consistently by the IndoBERT tokenizer and encoder [8]. Given an input sequence $x_{1:L}$, IndoBERT outputs token-level contextual representations $H = \{h_1, \dots, h_L\}$ (including the [CLS] token), which capture bidirectional context over text and nonverbal cues:

$$H = \{h_1, \dots, h_L\} = \text{IndoBERT}(x_{1:L}) \quad (2)$$

To reinforce sequential dependencies that may not be fully captured by a single pooled embedding, especially in short and noisy reviews, SIEmo-LSTM applies a BiLSTM on the IndoBERT token outputs. The forward and backward states are computed and concatenated as:

$$u_t^{\rightarrow} = \text{LSTM}_\text{f}(h_t, u_{t-1}^{\rightarrow}),\ u_t^{\leftarrow} = \text{LSTM}_\text{b}(h_t, u_{t+1}^{\leftarrow}) \quad (3)$$

$$u_t = [u_t^{\rightarrow}; u_t^{\leftarrow}] \quad (4)$$

A fixed-length representation is obtained via pooling and classified with a multiclass softmax over $C = 3$ sentiments:

$$z = \frac{1}{L}\sum_{t=1}^{L} \text{b}u_t,\ \hat{y} = \text{softmax}(Wz + b) \quad (5)$$

This design keeps the model end-to-end trainable while allowing BiLSTM gating to capture sequential patterns effectively [24], and is consistent with prior sentiment modeling practices that combine contextual encoders with BiLSTM-style refinements [25, 26].

G. Training Setting

Model training is conducted for a fixed number of epochs (as configured in the experiments) and monitored using the validation split. The overall training setup follows a typical fine-tuning paradigm where IndoBERT supplies strong contextual representations [8], while the BiLSTM and classifier layers specialize the model for multimodal sentiment inference.

III. RESULTS AND DISCUSSION

A. Evaluation Protocol and Metric

All SIEmo-LSTM variants were trained for 25 epochs and evaluated using Accuracy and Macro-F1. Macro-F1 better reflects balanced performance across negative, neutral, and positive classes under class imbalance. Macro-F1 is given by:

$$MacroF1 = \frac{F1_{neg} + F1_{neu} + F1_{pos}}{3} \quad (6)$$

B. Main Results

Table II reports results under expert labeling (Text=Expert, IE=Expert). Table III reports results under lexicon/MC labeling [Text=Lexicon, IE=Multi Converted (MC)]. Each table compares four modality configurations across three balancing settings (Original, ROS, RUS). The reported values are the testing Accuracy and Macro-F1 derived from the per-class F1 values.

TABLE II. MAIN RESULTS TEXT=EXPERT, SE=ESR, IE=EXPERT

Configuration	Original Acc	Original Macro-F1	ROS Acc	ROS Macro-F1	RUS Acc	RUS Macro-F1
Text Only	0.7708	0.6067	0.939	0.9367	0.7265	0.73
Text+SE	0.8458	0.6533	0.9637	0.9655	0.8125	0.8
Text+SE+IE	0.9407	0.6233	0.9935	0.9967	0.7143	0.6667
IE Only	0.8458	0.6537	0.9758	0.9767	0.8706	0.8633

TABLE III. MAIN RESULTS TEXT=LEXICON, SE=ESR, IE=MULTI CONVERTED

Configuration	Original Acc	Original Macro-F1	ROS Acc	ROS Macro-F1	RUS Acc	RUS Macro-F1
Text Only	0.6601	0.6267	0.7138	0.71	0.5843	0.58
Text+SE	0.747	0.7167	0.8593	0.86	0.8409	0.84
Text+SE+IE	0.8893	0.5867	0.9708	0.9733	0.6667	0.6433
IE Only	0.9763	0.9467	0.9867	0.9867	0.8732	0.87

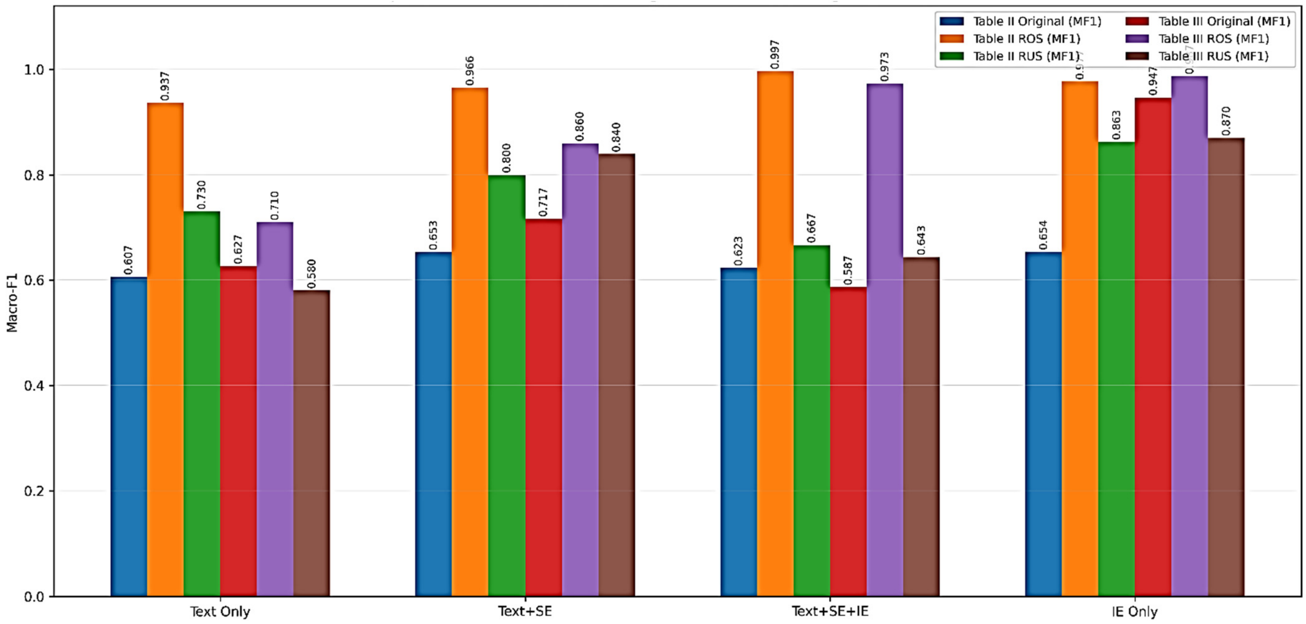

Fig. 3. Macro-F1 main results with Expert and MC.

Overall, the results show consistent incremental benefits from incorporating nonverbal cues. Under both labeling setups, adding Unicode emojis (Text+SE) improves over Text Only in the Original setting, indicating that SE provides complementary affective evidence. The full tri-modal Text+SE+IE becomes most reliable when imbalance is handled with ROS, achieving the highest Accuracy and Macro-F1 in both Tables II and III. In contrast, RUS often degrades performance, especially for the tri-modal setting, suggesting that undersampling removes informative variability that is important for noisy short reviews. To improve readability, Figure 3 visualizes Macro-F1 across configurations and balancing strategies for both labeling setups.

C. Discussion

Across both labeling setups, Unicode emoji cues (SE) consistently improve performance over text-only modeling. In the Original setting, Text+SE outperforms Text Only in Table II (0.8458 vs 0.7708 Accuracy) and Table III (0.7470 vs 0.6601 Accuracy), and it also yields higher Macro-F1 under ROS (Table II: 0.9655 vs 0.9367; Table III: 0.8600 vs 0.7100). When class imbalance is addressed, the tri-modal configuration (Text+SE+IE) achieves the strongest results under ROS (Table II: 0.9935 Accuracy / 0.9967 Macro-F1; Table III: 0.9708 / 0.9733), showing that ASCII emoticons/icon emojis (IE) add polarity cues beyond text and SE. In contrast, the Original condition shows high Accuracy but low Macro-F1 (Table II: 0.9407 vs 0.6233; Table III: 0.8893 vs 0.5867), indicating errors concentrated in minority classes—especially neutral, which is ambiguous and prone to being absorbed into polar predictions.

The balancing strategy is decisive: ROS improves both Accuracy and Macro-F1 across configurations, whereas RUS often degrades performance, particularly for Text+SE+IE (Table II: 0.7143 / 0.6667; Table III: 0.6667 / 0.6433). This suggests that oversampling increases minority exposure while preserving training diversity, whereas undersampling removes informative variation in noisy reviews. Although IE-only can be strong in Table III (Original: 0.9763 / 0.9467; ROS: 0.9867 / 0.9867), the overall evidence supports tri-modal fusion with ROS as the most reliable choice for balanced performance on Indonesian reviews (Tables II and III). These findings are consistent with prior works showing that emoji-text integration improves sentiment prediction via fusion or emoji-aware representations [14-17], and that emoticon-aware pipelines yield gains when emoticons are blended with text or injected as explicit affective features [18, 19]. In [6], strong performance was reported on noisy Indonesian comments when robust preprocessing was paired with sequential modeling.

Table IV compares the results of this study with representative emoji-aware sentiment studies that are mostly bimodal (text+SE) [14, 16, 17]. Although the cross-study metrics are not directly comparable, Table IV highlights that prior systems typically omit ASCII emoticons (IE), whereas SIEmo-LSTM models text+SE+IE. BERT+BiLSTM+ROS achieves the highest reported accuracy (0.99), supporting the benefit of explicit emoticon modeling and imbalance-aware training for Indonesian reviews.

TABLE IV. COMPARISON WITH RELATED STUDIES

Ref.	language	Text	SE	IE	Method	Acc
[14]	Arabic	√	√	x	Machine learning, CBOW, Skip Gram	0.84
[16]	China	√	√	x	Deep learning, Bi-LSTM	0.81
[17]	English	√	√	x	Deep learning, CNN, Bi-LSTM	0.73
Proposed	**Indonesia**	√	√	√	**Deep learning, BERT, Bi-LSTM, ROS**	**0.99**

IV. CONCLUSIONS

This paper presented SIEmo-LSTM, a tri-modal framework that integrates Indonesian text (T), Unicode emojis (SE), and ASCII emoticons/icon emojis (IE). Across labeling setups, multimodal cues outperform text-only baselines and require imbalance handling for macro-level robustness. Under Expert/Expert, Text+SE+IE+ROS achieves 0.9935 Accuracy and 0.9967 Macro-F1, exceeding Text Only+ROS (0.9390/0.9367) and Text+SE+ROS (0.9637/0.9655). Under Lexicon/MC, it reaches 0.9708/0.9733, surpassing Text+SE+ROS (0.8593/0.8600) and Text Only+ROS (0.7138/0.7100). In contrast, RUS degrades performance (Text+SE+IE: 0.7143/0.6667 and 0.6667/0.6433), indicating that undersampling discards informative variability in noisy reviews. In general, explicitly modeling ASCII emoticons alongside Unicode emojis yields more reliable, class-balanced Indonesian sentiment recognition—especially with ROS. Many prior emoji-aware models remain bimodal. Although cross-study metrics are not directly comparable, the proposed tri-modal BERT+BiLSTM+ROS setting attains the highest reported accuracy in the comparison (0.99), supporting ASCII emoticons as a first-class modality.

Future work includes releasing a human-annotated tri-modal benchmark, cross-domain evaluation, improved context-aware emoticon disambiguation, and learned fusion/calibration (especially for the neutral class) to reduce neutral-to-polar collapse.

ACKNOWLEDGMENT

This research is supported by Universitas Negeri Malang and Universitas Nurul Jadid.

REFERENCES

[1] B. Andrian, T. Simanungkalit, I. Budi, and A. F. Wicaksono, "Sentiment Analysis on Customer Satisfaction of Digital Banking in Indonesia," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 3, 2022, https://doi.org/10.14569/IJACSA.2022.0130356.

[2] G. Buntoro, R. Arifin, G. Syaifuddiin, A. Selamat, O. Krejcar, and F. Hamido, "The Implementation of the Machine Learning Algorithm for the Sentiment Analysis of Indonesia's 2019 Presidential Election," *IIUM Engineering Journal*, vol. 22, no. 1, pp. 78–92, Jan. 2021, https://doi.org/10.31436/iiumej.v22i1.1532.

[3] F. Kurniawan, B. Badruddin, and P. A. Wibawa, "Identification of islamophobia sentiment analysis on Twitter using text mining language detection," *Journal of Positive School Psychology*, vol. 6, no. 5, pp. 8286–8294, 2022.

[4] N. Risa, D. D. Prasetya, W. N. Hidayat, P. I. Maula, I. M. Wirawan, and S. Y. Setiawan, "Sentiment Analysis of 'Kampus Merdeka' on Twitter Using Support Vector Machine (SVM) Algorithm," in *2024 IEEE 2nd International Conference on Electrical Engineering, Computer and Information Technology (ICEECIT)*, Nov. 2024, pp. 163–168, https://doi.org/10.1109/ICEECIT63698.2024.10859978.

[5] R. Kusumaningrum, I. Z. Nisa, R. Jayanto, R. P. Nawangsari, and A. Wibowo, "Deep learning-based application for multilevel sentiment analysis of Indonesian hotel reviews," *Heliyon*, vol. 9, no. 6, June 2023, Art. no. e17147, https://doi.org/10.1016/j.heliyon.2023.e17147.

[6] D. Ariyus, D. Manongga, and I. Sembiring, "Enhancing Sentiment Analysis of Indonesian Tourism Video Content Commentary on TikTok: A FastText and Bi-LSTM Approach," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 18020–18028, Dec. 2024, https://doi.org/10.48084/etasr.8859.

[7] S. Supriyono, "Analyzing Audience Sentiments in Digital Comedy: A Study of YouTube Comments Using LSTM Models," *Journal of Applied Data Sciences*, vol. 5, no. 4, pp. 1877–1889, Dec. 2024, https://doi.org/10.47738/jads.v5i4.393.

[8] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North*, 2019, pp. 4171–4186, https://doi.org/10.18653/v1/N19-1423.

[9] I. Darmawan, H. Elmunsyah, and D. D. Prasetya, "ALBERTIR: A BERT-Based Pretraining for Indonesian Religious Texts Using Qur'an and Hadith Translations," *Engineering, Technology & Applied Science Research*, vol. 15, no. 5, pp. 28307–28312, Oct. 2025, https://doi.org/10.48084/etasr.12977.

[10] L. B. Ilmawan, M. Muladi, and D. D. Prasetya, "Negation handling for sentiment analysis task: approaches and performance analysis," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 3, June 2024, Art. no. 3382, https://doi.org/10.11591/ijece.v14i3.pp3382-3393.

[11] L. Li and X. T. Wang, "Nonverbal communication with emojis in social media: dissociating hedonic intensity from frequency," *Language Resources and Evaluation*, vol. 57, no. 1, pp. 323–342, Mar. 2023, https://doi.org/10.1007/s10579-022-09611-6.

[12] P. Kralj Novak, J. Smailović, B. Sluban, and I. Mozetič, "Sentiment of Emojis," *PLOS ONE*, vol. 10, no. 12, Dec. 2015, Art. no. e0144296, https://doi.org/10.1371/journal.pone.0144296.

[13] D. Rodrigues, M. Prada, R. Gaspar, M. V. Garrido, and D. Lopes, "Lisbon Emoji and Emoticon Database (LEED): Norms for emoji and emoticons in seven evaluative dimensions," *Behavior Research Methods*, vol. 50, no. 1, pp. 392–405, Feb. 2018, https://doi.org/10.3758/s13428-017-0878-6.

[14] S. Al-Azani and E. S. M. El-Alfy, "Early and Late Fusion of Emojis and Text to Enhance Opinion Mining," *IEEE Access*, vol. 9, pp. 121031–121045, 2021, https://doi.org/10.1109/ACCESS.2021.3108502.

[15] S. Velampalli, C. Muniyappa, and A. Saxena, "Performance Evaluation of Sentiment Analysis on Text and Emoji Data Using End-to-End, Transfer Learning, Distributed and Explainable AI Models," *Journal of Advances in Information Technology*, vol. 13, no. 2, 2022, https://doi.org/10.12720/jait.13.2.167-172.

[16] C. Liu *et al.*, "Improving sentiment analysis accuracy with emoji embedding," *Journal of Safety Science and Resilience*, vol. 2, no. 4, pp. 246–252, Dec. 2021, https://doi.org/10.1016/j.jnlssr.2021.10.003.

[17] S. Kusal, S. Patil, and K. Kotecha, "Multimodal text-emoji fusion using deep neural networks for text-based emotion detection in online communication," *Journal of Big Data*, vol. 12, no. 1, Feb. 2025, Art. no. 32, https://doi.org/10.1186/s40537-025-01062-4.

[18] H. Zou and K. Xiang, "Sentiment Classification Method Based on Blending of Emoticons and Short Texts," *Entropy*, vol. 24, no. 3, Mar. 2022, Art. no. 398, https://doi.org/10.3390/e24030398.

[19] Y. J. Su, C. H. Chen, T. Y. Chen, and C. C. Cheng, "Chinese Microblog Sentiment Analysis by Adding Emoticons to Attention-Based CNN," *Journal of Internet Technology*, vol. 21, no. 3, pp. 821–829, May 2020.

[20] Y. Chen, J. Yuan, Q. You, and J. Luo, "Twitter Sentiment Analysis via Bi-sense Emoji Embedding and Attention-based LSTM," in *Proceedings of the 26th ACM International Conference on Multimedia*, Oct. 2018, pp. 117–125, https://doi.org/10.1145/3240508.3240533.

[21] Y. Wang, "Iteration-based naive Bayes sentiment classification of microblog multimedia posts considering emoticon attributes," *Multimedia Tools and Applications*, vol. 79, no. 27–28, pp. 19151–19166, July 2020, https://doi.org/10.1007/s11042-020-08797-7.

[22] M. A. Ullah, S. M. Marium, S. A. Begum, and N. S. Dipa, "An algorithm and method for sentiment analysis using the text and emoticon," *ICT Express*, vol. 6, no. 4, pp. 357–360, Dec. 2020, https://doi.org/10.1016/j.icte.2020.07.003.

[23] F. Koto and G. Y. Rahmaningtyas, "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," in *2017 International Conference on Asian Language Processing (IALP)*, Dec. 2017, pp. 391–394, https://doi.org/10.1109/IALP.2017.8300625.

[24] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5929–5955, Dec. 2020, https://doi.org/10.1007/s10462-020-09838-1.

[25] C. Suman, S. Saha, P. Bhattacharyya, and R. S. Chaudhari, "Emoji Helps! A Multi-modal Siamese Architecture for Tweet User Verification," *Cognitive Computation*, vol. 13, no. 2, pp. 261–276, Mar. 2021, https://doi.org/10.1007/s12559-020-09715-7.

[26] M. Fernández-Gavilanes, E. Costa-Montenegro, S. García-Méndez, F. J. González-Castaño, and J. Juncal-Martínez, "Evaluation of online emoji description resources for sentiment analysis purposes," *Expert Systems with Applications*, vol. 184, Dec. 2021, Art. no. 115279, https://doi.org/10.1016/j.eswa.2021.115279.