Jurnal Informatika Universitas Pamulang **ISSN: 2541-1004**
Penerbit: Fakultas Ilmu Komputer Universitas Pamulang **e-ISSN: 2622-4615**
Vol. 11, No. 2, June 2026 (111-122) **https://doi.org/10.32493/jiup.v11i2.58892**

Early Detection of Depression in Indonesian Social Media Text with IndoBERT

Ninda Khoirunnisa [a,1*], Sheraton Pawestri [a,2]

[a] Department of Informatics, Faculty of Industrial Technology, Ahmad Dahlan University, Yogyakarta, Indonesia
e-mail: [1]ninda@tif.uad.ac.id, [2]sheraton@tif.uad.ac.id
*Corresponding author

Submitted Date: March 25th, 2026 Reviewed Date: April 10th, 2026
Revised Date: April 30th, 2026 Accepted Date: June 20th, 2026

Abstract

Depression has led to many cases of suicide across the world. Hence, early screenings are needed to provide the right medical treatments. Recent studies have attempted to use social media text which have become an integral part of people's daily lives to early detect depression, since users primarily post and express their thoughts in the form of text or media that reflect their current mental state. This study aims to fine-tune the pre-trained IndoBERT model on labeled Indonesian social media texts from X (previously Twitter) for binary depression-related text classification. The dataset used in this study, "Depression and Anxiety in Twitter (ID)," consists of 2,201 labeled texts on train set and 550 labeled texts on test set. Several evaluation criteria such as recall, precision, F1-score, and accuracy were used on an unseen test set to calculate how well the model performed. The results show that the IndoBERT model achieved an overall accuracy of 86% and a 74.75% F1-score in identifying the depressed class, notably achieving a strong precision score of 79.17%. These findings show the potential of IndoBERT fine-tuning for Indonesian text-based depression-related classification while highlighting the need for further improvements and validation.

Keywords: Depression; Fine-Tuning; IndoBERT; Text-Classification; Twitter

1. Introduction

Depression is a major mental health problem that affect individuals' daily lives and overall well-being significantly. According to the 2023 Indonesian Health Survey (*Survei Kesehatan Indonesia*/SKI), the prevalence of depression among Indonesians aged 15 years and older was 1.4% [1]. The complicated nature of mental disorders often makes it difficult for psychiatrists to identify symptoms early enough to provide appropriate treatment [2], [3].

In modern days, social media has formed a digital space where people frequently express their feelings, including deep sadness, emotional shifts, and behavioral changes [3]. Social medias such as Twitter (now X) and Reddit provide sources of user-generated text in which users may express emotions and experiences related to psychological well-being [4], [5]. The availability of such data provides an opportunity to explore Artificial Intelligence (AI) based approaches for analyzing depression-related content that may support preliminary screening [6].

Despite its potential, detecting depressive symptoms from social media texts remains challenging because such data are often unstructured, noisy, and often unreliable [7]. Additionally, tweets from social media are characterized by informal vocabulary, abbreviations, emojis, and hashtags, which pose challenges for NLP tasks [8]. This challenge is relevant in the Indonesian context, where social media users frequently use informal language, slang, and abbreviations, making it challenging to accurately classify text containing signs of emotional distress [9]. These characteristics can limit the performance of traditional machine learning algorithms that rely on basic feature extraction

hmethods such as TF-IDF or Bag-of-Words. Such methods may not capture the deeper semantic and contextual information needed to distinguish genuine sign of depression from ordinary expression of sadness [5], [9]. Within the domain of Natural Language Processing (NLP), te architecture known as

Bidirectional Encoder Representations from Transformers (BERT) has became state-of-the-art in text classification due to its capability in capturing complex sequential dependencies and rich context [3], [7], [10]. Previous studies have shown that BERT and its variants achieved competitive or higher performance that traditional deep learning architectures in depression and suicide detection tasks on social media [5], [10].

For text specifically written in Indonesian, localized Transformer models have been developed to address the unique linguistic challenges of the region. IndoBERT, a state-of-the-art language model pre-trained on a massive corpus of Indonesian text, has been applied to Indonesian NLP tasks involving social media text [11]. Recent studies have reported promising performance from fine-tuned pretrained models such as IndoBERT for classifying emotions linked to depression in Indonesian tweets, bridging the gap between complex NLP architectures and local language nuances [11], [12]. However, further evaluation is still needed to assess the effectiveness of IndoBERT for depression-related text classification tasks on Indonesian social media text, particularly using publicly available Indonesian datasets.

Recent studies have demonstrated the potential of transformer-based models for depression-related text classification in Indonesian social media. Study [13] fine-tuned IndoBERT in classifying potentionally depressive text on X. The dataset used in this study consists of 5,000 Indonesian tweets that and the model achieved an accuracy, precision, recall, and F1-score of 87%. Similarly, [14] explored multiple pretrained transformer models, including IndoBERT, RoBERTa, and DistilBERT combined with TF-IDF and BoW feature representations to classify depression-related emotions in Indonesian tweets using a dataset of 5,271 tweets. The combination of DistilBERT and TF-IDF achieved the highest performance among the evaluated configurations, with an F1-score of 0.699 and precision of 0.714. Another study [15] evaluated the effectiveness of IndoBERT in identifying potentially depressive text from social media using a dataset of 37,554 textx from X, achieving an accuracy, precision, recall, and F1-score of 94,91%.

However, existing studies have primarily focused on evaluating transformer-based models on relatively larger datasets, while the performance of fine-tuned IndoBERT under a relatively smaller labeled-data setting and in comparison with a task-specific deep elarning architecture remains less explored. Therefore, this study evaluates fine-tuned IndoBERT which uses pretrained contextual representations and compares it with a CNN+BiLSTM baseline which uses the IndoBERT tokenizer followed by a task-specific trainable embedding layer and CNN+BiLSTM feature extraction using the "Depression and Anxiety in Twitter (ID)" dataset. Thie dataset is used to examine the models under a relatively small labeled-data setting, with preprocessing consisting of text cleaning, case folding, normalization, and IndoBERT tokenization. The contribution of this study is a comparative evaluation of pretrained contextual representations and task-specific trainable representations for depression-related text classification in Indonesian social media text, providing insight into the performance of a pretrained language model when labeled data are relatively limitedy.

2. Research Methodology

a. Dataset

The dataset used in our study consists of Twitter/X posts which serve as the corpus for depressive and non-depressive texts, sourced from Kaggle [16].

This dataset consists of 3 files: datd_train, datd_test, and datd_rand. However, only datd_train and datd_test were used in this study because datd_rand contains overlapping positive samples from datd_test, potentially causing data leakage. All files in this dataset have two columns, namely 'text' that contain the unstructured Twitter/X posts and 'label' that contain the label (1/0) for the texts. The '1' label is for Twitter posts classified as depressed text, meanwhile '0' shows otherwise. Table 1 shows the example of instances with 1 and 0 label, meanwhile Table 2 shows the distribution of two labels across all files. This study utilizes datd_train for training and validation purpose, and datd_test for testing purpose.

Table 1. Example of instances from the dataset

Text (Indonesian)	Text (English)	Label
Udah ngalihin fokus ke yang lain tapi tetep cemas	Already diverted focus to other things but I'm still anxious	1
wkwkwk apaansi lawak	wkwkwk so funny	0

ISSN: 2541-1004
e-ISSN: 2622-4615
https://doi.org/10.32493/jiup.v11i2.58892

Table 2. Label distribution across train, test, and rand files

Files	Label	
	0	1
datd_train	1468	733
datd_test	389	161
datd_rand	6247	733

b. Methodology

Figure 1 shows the methodology of this research. The tweet text undergoes a sequence of processing steps designed to remove noise and keep semantically relevant information for transformer-based models such as IndoBERT. Transformer models exploit contextual embeddings, and thus preprocessing must avoid discarding meaningful syntactic or semantic features.

First, case folding converts all characters into lowercase. This is appropriate when using an uncased pretrained language model such as indobenchmark/indobert-base-p1, which expects lowercase input tokens. Lowercasing standardised text and reduces vocabulary sparsity by eliminating capitalization differences [17].

Second, emojis are important and commonly used in social media and information that is valuable for classification. Transforming emojis into semantic token labels (😂 →face_with_tears_of_joy) allows models to learn associations between emotional expressions and the target labels. Previous research demonstrates that emojis contain sentiment information and should be treated as meaningful tokens rather than discarded [18].

Third, noises in social media text such as URLs, user mentions (@_), and hashtags (#) was cleaned using regular expressions. Fourth, social media users frequently elongate words for emphasis by repeating characters (e.g., 'capekkk'). A normalization function is applied to remove consecutive identical characters to a maximum of two (e.g., 'capekk'). This prevents the tokenizer from treating elongated variations of the same word as unknown tokens.

Additionally, tweets frequently contain informal language and slang abbreviations. Studies on Indonesian Twitter dataset emphasize the importance of normalization for classification task such as sentiment analysis [19]. Without normalization, such nonstandard tokens increase vocabulary sparsity and affect model learning. In this study, slang normalization was used to converts variants such as 'gk' or 'bgt' into standard Indonesian equivalents ('tidak'/'not', 'banget'/'very'), therefore preventing the token to be considered out-of-vocabulary.

Finally, text from social media such as Twitter/X often contain irregular spacing. Replacing multiple whitespace characters with single spaces helps keeping the text consistent before tokenization. This preprocessing step ensures that the input text is well-structured for subsequent tokenization [8].

After preprocessing, the 2,201 samples from datd_train file were splitted into training and validation sets using an 80:20 ratio. While the training set was used to fine-tune the IndoBERT model, the validation set was used to evaluate the model's accuracy and loss metrics.

IndoBERT [20] represents an advanced pre-trained language model designed specifically to the Indonesian text. Its structural framework mirrors that of the Bidirectional Encoder Representations from Transformer (BERT) [21], [22]. Consequently, the system evaluates the entire word sequences at once, capturing deep contextual relationships from both directions. The key part of this architecture is the self-attention mechanism, which was introduced in the original Transformer model [23]. Self-attention allows the model to determine how much attention should be given to each word based on the other words in the same semtence. The attention mechanism uses three matrices, namely queries (Q), keys (K), and values (V) with the attention scores scaled according to the dimensionality of the key (d_k).

ISSN: 2541-1004
e-ISSN: 2622-4615
https://doi.org/10.32493/jiup.v11i2.58892

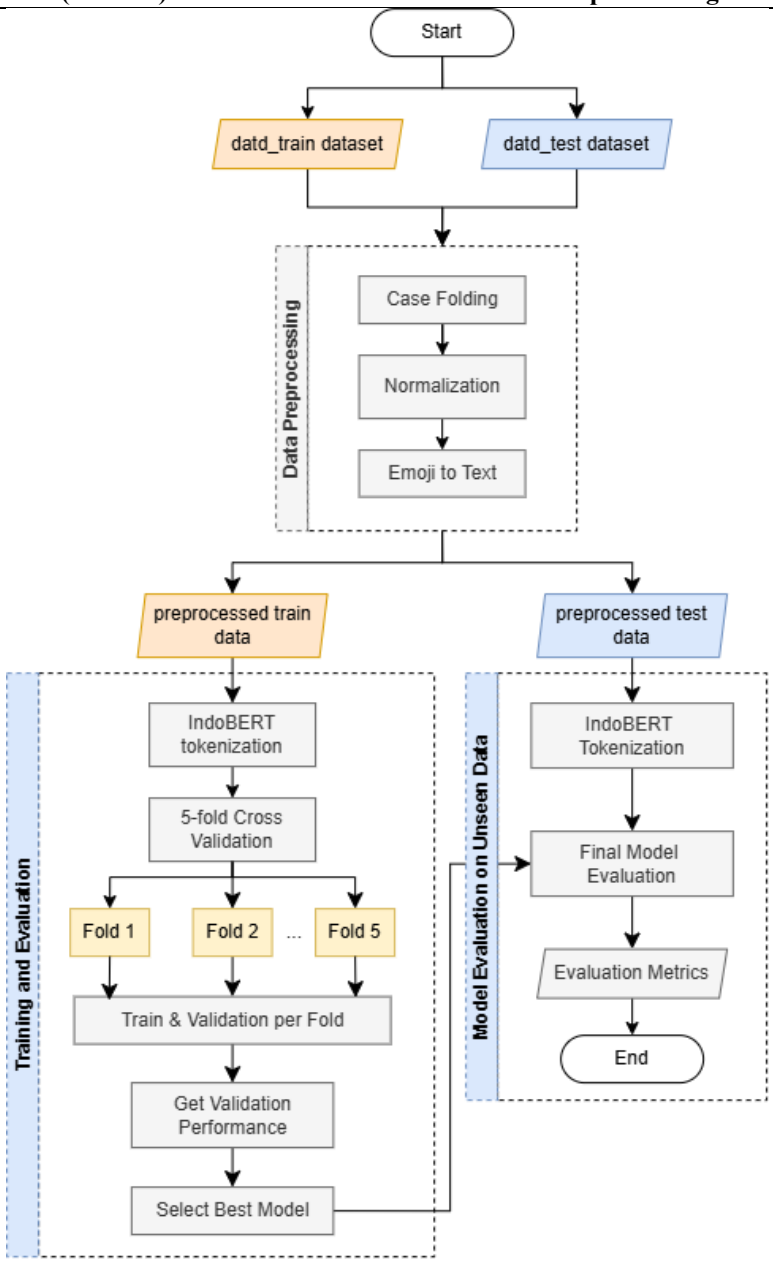

Figure 1. Research methodology

IndoBERT was pretrained using Indo4B dataset, a large Indonesian text corpus containing approximately 4 billion words collected from various sources, including news articles, Wikipedia, and social media data [22]. This pre-training allows the model to learn Indonesian language pattern, including syntactic rules, semantic meanings, and contextual slang commonly used in Indonesian.

Utilizing the baseline IndoBERT architecture for a specialized NLP task such as depression-related text classification involves a process called fine-tuning [24], [25]. Before fine-tuning the IndoBERT with the training set, the data was tokenized using pre-trained IndoBERT's WordPiece tokenizer. Tokenization breaks down the normalized tweet into subword units, mapping them to specific integer IDs based on vocabulary built from the model. To denote the start and end of a tweet sequence, a [CLS] token marks the start of the input, while a [SEP] token indicates its conclusion. The model maps these tokens to their corresponding embeddings, combining them with segment embeddings and positional embeddings to retain the sequence order. Because transformer architecture requires input sequence of uniform length, the tokenized sequences are padded or truncated to a pre-determined maximum sequence length. During this step the system also creates an attention mask to differentiate actual word tokens from padding, preventing the model from processing the padded sections during training [26].

Following tokenization, the baseline IndoBERT architecture is fine-tuned to the specific tasks. The entire sequence of embeddings is fed into the stacked Transformer encoder layers [27]. By utilizing feed-forward networks alongside multi-head self-attention mechanisms, the system dynamically adjusts the representation of each token based on its context.

For sequence classification, the final hidden state of the special [CLS] token is extracted, as it serves as the aggregate semantic representation of the entire tweet. A dedicated classification head consisting of a linear layer is added above this core IndoBERT structure. The classification head produces raw, unnormalized prediction scores known as logits. To determine the final discrete class prediction ($\hat{y}$), an argmax function is utilized to select the class associated with the highest logit value, following Equation 1.

$$\hat{y} = argmax\left(h_{[CLS]}W^T + b\right) \quad (1)$$

During fine-tuning, all parameters of the IndoBERT model, along with the newly added classification layer, are trained jointly using the labeled depression dataset. The training loss is computed using Cross-Entropy Loss which internally applies a softmax function to the logits to calculate the predicted probability distribution before computing the error. The model weights were updated using the AdamW optimizer to minimize this loss. During this process, IndoBERT was able to adapt its pretrained language representations to the characteristics of the training dataset [25].

Several hyperparameters were also set for the fine-tuning process to ensure reproducibility. The configuration is shown in Table 3.

Table 3. The Hyperparameter Settings

Hyperparameter	Value
Model name	Indobenchmark/indobert-base-p1
Batch size	16
Dropout	0.3
Maximum epochs	10
Learning rate (LR)	5e-6
Maximum sequence length	128
Early stopping patience	2

During training and model selection, the 5-fold stratified cross validation was conducted only on the training set (datd_train), with each fold using 80% of the data for training and 20% for validation. At each epoch, the validation accuracy and loss were monitored to track the model's learning. Early stopping was applied with a patience = 2, meaning that training was terminated if the validation loss did not improve for two consecutive epoch.

For each fold, the checkpoint with the highest validation accuracy achieved during training was first identified. The best checkpoints from each fold were then compared based on their highest validation accuracy, and the checkpoint with the highest validation accuracy across the 5 folds was selected as the best model. This selected model was then evaluated on the independent test set (datd_test) using accuracy, precision, recall, and F1-Score. Thus, the test set remained unseen throughout the model training and selection. The calculations for each metric are represented in Equations 2-5, respectively. In the context of identifying depressive social media text, the F1-Score (defined in Equation 5) was used as the main metrics or evaluators as it calculates the harmonic mean between recall and precision. Additionally, this metric is preferred compared to accuracy which may not fully reflect the model's performance, especially on evaluating model trained on imbalances datasets used in this research.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

ISSN: 2541-1004
e-ISSN: 2622-4615
https://doi.org/10.32493/jiup.v11i2.58892

Where:

- TP: True positive (tweet indicating depression predicted as indicating depression)
- TN: True negative (tweet not indicating depression predicted as not indicating depression)
- FP: False positive (tweet not indicating depression falsely predicted as indicating depression
- FN: False negative (tweet indicating depression falsely predicted as not indicating depression)

3. Result and Discussion

The fine-tuning of IndoBERT for classifying depressive text in Indonesian social media text was conducted using 5-fold stratified cross validation over a maximum of 10 epochs with early stopping applied using a patience of 2. The model's learning progression was monitored through training and validation loss and accuracy as shown in Figure 2.

The implementation of patience = 2 during training forced the training process to be stopped when an improvement in validation loss for two consecutive epochs was recorded. As shown in Figure 2, the validation loss initially decreased across all folds. However, the validation loss began to increase after reaching a minimum value, which indicates an overfitting.

Across 5 folds, the minimum validation loss value was achieved before the training process reached the last epoch. In Fold 1, the minimum validation loss value of 0.3793 was reached on epoch 4. A similar pattern was found in Fold 2, where the minimum validation loss of 0.4462 was achieved at epoch 4. Similarly, the lowest validation loss of 0.4452 was recorded at epoch 3 in Fold 3, a minimum of 0.4523 at epoch 3 in Fold 4, and a minimum of 0.4668 at epoch 3 in Fold 5. As the validation loss increased for the next two consecutive epochs across all folds, the training was stopped to prevent overfitting. Following the 5-fold stratified cross-validation, the model checkpoint from the fold with the highest validation accuracy was selected as the best model. This strategy ensures that the chosen model represents the best-performing configuration among the evaluated folds while maintaining generalization capability on unseen validation data used in this study.

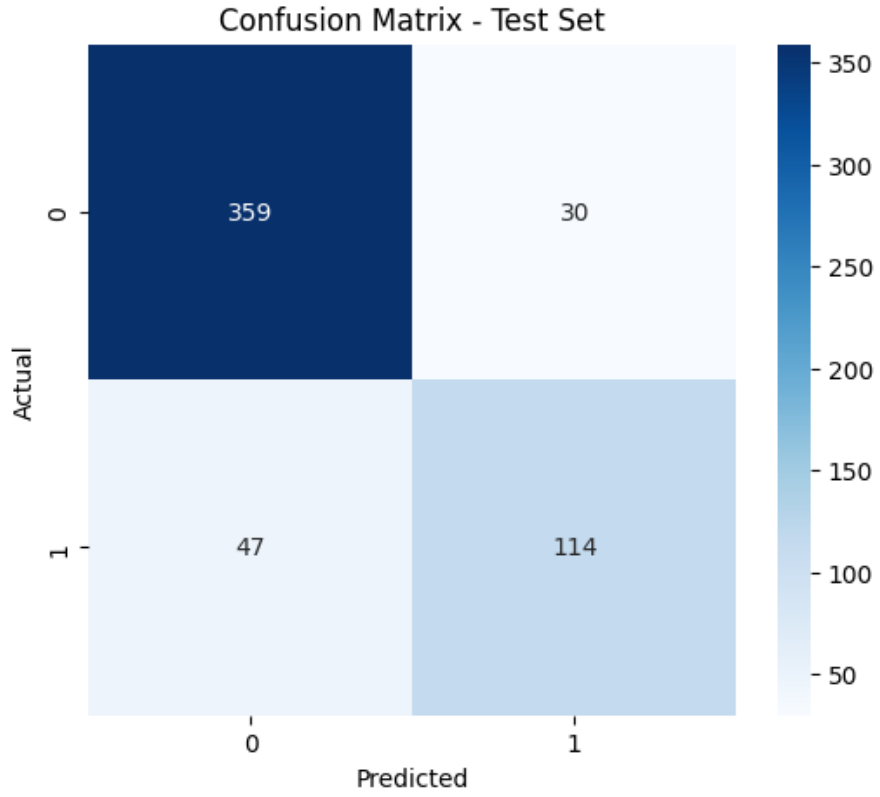

Figure 2. Confusion Matrix of IndoBERT

Jurnal Informatika Universitas Pamulang **ISSN: 2541-1004**
Penerbit: Fakultas Ilmu Komputer Universitas Pamulang **e-ISSN: 2622-4615**
Vol. 11, No. 2, June 2026 (111-122) **https://doi.org/10.32493/jiup.v11i2.58892**

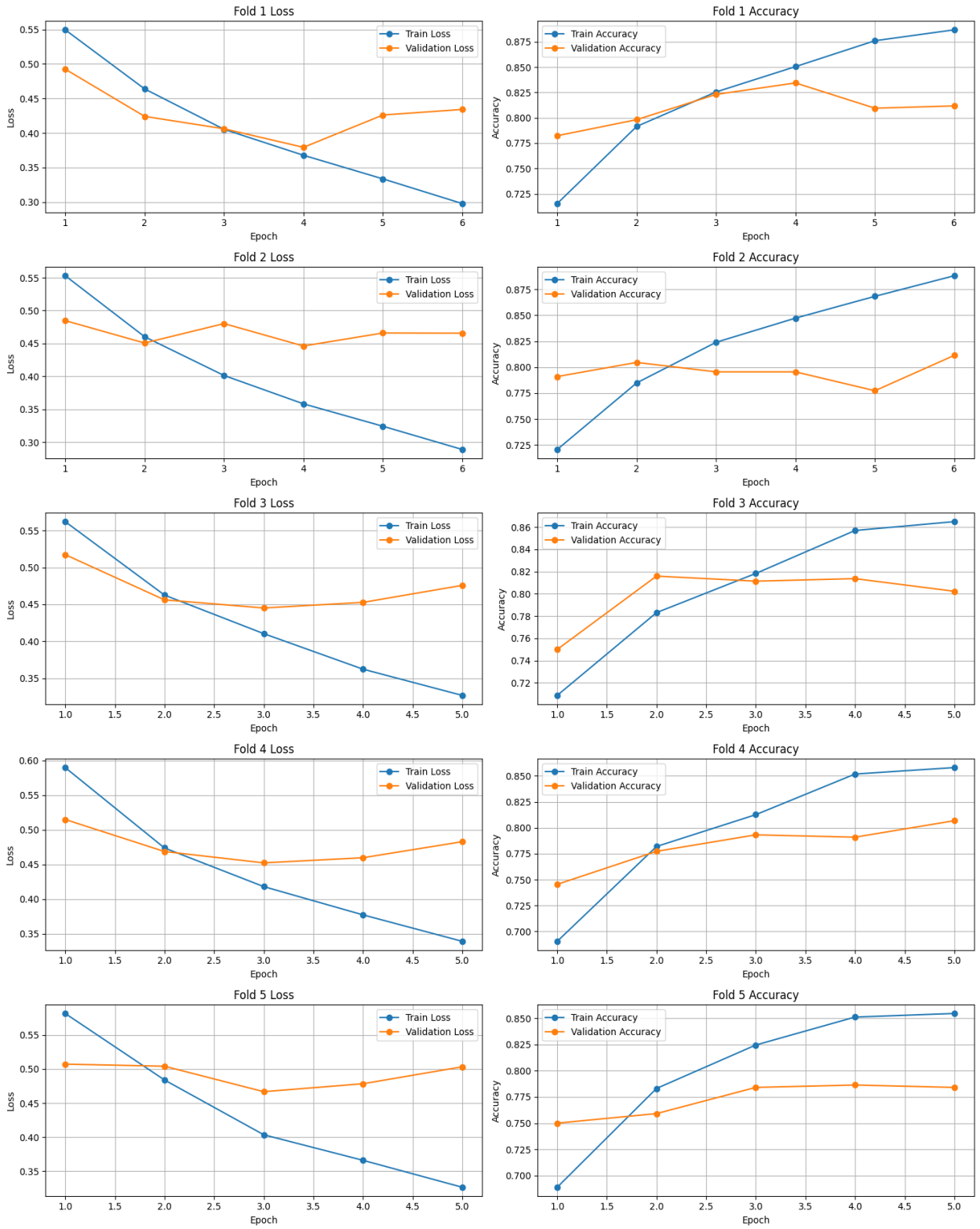

Figure 3. Loss and Accuracy per Epoch

The selected model was evaluated on a test set (datd_test) consisting of 550 unseen samples during training or validation. As shown in the confusion matrix (Figure 3), the model correctly classified 359 non-

depression instances (True Negatives) and 114 depression-related instances (True Positives). Meanwhile, 30 non-depression samples were incorrectly classified as depression (False Positives), and 47 depression samples were misclassified as non-depression (False Negatives). In total, 473 out of 550 test samples were classified correctly, giving the model an accuracy of 86.00%.

The classification report in Figure 4 shows the effectiveness of fine-tuned IndoBERT for both classes. For the non-depressive-text class (Class 0), the model obtained a precision of 88.42%, recall of 92.29%, and F1-score of 90.31%. For the depressive-text class (Class 1), the model achieved a precision of 79.17%, recall of 70.81%, and F1-score of 74.75%. The higher recall and precision score for Class 0 indicates that the model performed better in identifying non-depressive-text, while its performance on depressive-text was comparatively lower.

•••	precision	recall	f1-score	support
0	0.8842	0.9229	0.9031	389
1	0.7917	0.7081	0.7475	161
accuracy			0.8600	550
macro avg	0.8380	0.8155	0.8253	550
weighted avg	0.8571	0.8600	0.8576	550

Figure 2. Classification Report of IndoBERT Model

Furthermore, the differences in performance between the two classes in precision, recall, and F1-score may be related to the class distribution in the dataset, where the number of non-depressive-text samples (n = 1468) exceeded the number of depressive-text samples (n = 733). Consequently, the model was exposed to a larger number of non-depression examples during training.

To provide a comparative baseline, the performance of fine-tuned IndoBERT model was compared to a CNN+BiLSTM architecture (Table 4), which was trained and evaluated on the same dataset. The results showed that IndoBERT achieved higher precision, recall, F1-score, and accuracy than CNN+BiLSTM on the test set, as shown in Figure 5 and Table 5. Additionally, Figure 6 shows the learning progression of the CNN+BiLSTM model over a maximum of 10 epochs. Early stopping was triggered at epoch 4 because the validation loss did not improve for two consecutive epochs, despite the decrease on training loss and increase on training accuracy

On the test set, IndoBERT correctly identified 114 depressive-text samples, whereas CNN+BiLSTM identified 110. In addition, CNN+BiLSTM produced 51 false-negative predictions compared to 47 generated by IndoBERT. These results indicate the differences in the models' ability to identify depressive-text within the test set.

Table 4. The Hyperparameter Settings of CNN+BiLSTM

Hyperparameter	Value
Vocab size	IndoBERT Tokenizer's vocab size
Embedding dimension	200
Number of onvolutional filters	100
Learning rate (LR)	1e-3
Maximum epochs	10 (patience = 2)
Number of LSTM layers	2
Dropout rate	0.3
LSTM hidden size	128
Kernel size	3, 4, 5

ISSN: 2541-1004
e-ISSN: 2622-4615
https://doi.org/10.32493/jiup.v11i2.58892

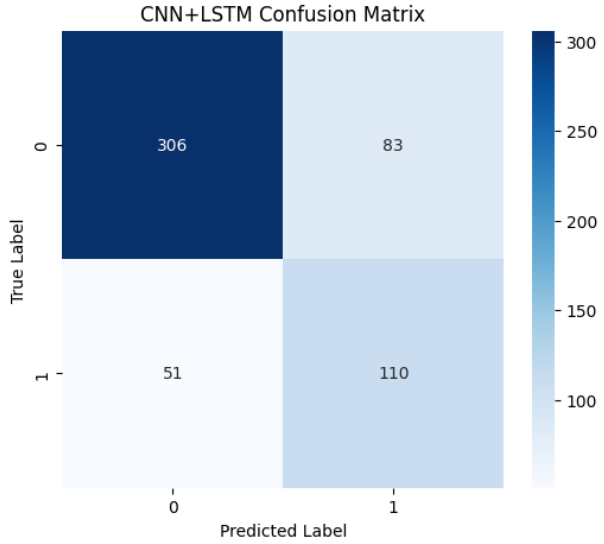

Figure 3. Classification Report of CNN+BiLSTM Model

Table 5. Comparison of IndoBERT and CNN+BiLSTM

Metrics	IndoBERT	CNN+BiLSTM
Accuracy	**86%**	75.64%
Precision – 0 class (non-depressive-text)	**88.42%**	85.71%
Precision – 1 class (depressive-text)	**79.17%**	56.99%
Recall – 0 class (non-depressive-text)	**92.29%**	78.66%
Recall – 1 class (depressive-text)	**70.81%**	68.32%
F1-Score – 0 class (non-depressive-text)	**90.31%**	82.04%
F1-Score – 1 class (depressive-text)	**74.75%**	62.15%

Figure 4. Training and Validation Accuracy Comparison of IndoBERT and CNN+BiLSTM

The observed differences may be related to the representation learning mechanisms used by the two approaches. CNN+BiLSTM extract features from a sequence of trainable word representations, while IndoBERT used contextualized representations learned during pretraining on large-scale Indonesian text corpora. This may allow IndoBERT to capture semantic information that may be difficult to capture for CNN+BiLSTM's feature extractor approach.

This result show that the fine-tuned IndoBERT model achieved better classification performance than the CNN+BiLSTM baseline on the evaluated test set. IndoBERT produced fewer false-negative prediction on the evaluated test set, indicating that IndoBERT correctly classified more depression-related samples correctly than the CNN+BiLSTM baseline in the test set. However, neither model classified all posts correctly. This shows that classifying depression-related social media remains challenging.

4. Conclusion

This study aimed to evaluate the performance of IndoBERT for binary depression-related text classification on Indonesian social media text using the Depression and Anxiety in Twitter (ID) dataset. To achieve this objective, Transformer-based token encoding was applied prior to model fine-tuning. The results indicate that the result of IndoBERT fine-tuning was able to classify depressive and non-depressive content, with an accuracy of 86%, a macro-average F1-score of 82.53%, and a weighted-average F1-score of 85.76% on the test set consisting of 550 data. These findings demonstrate that the research objective of evaluating IndoBERT for depression-related text classification within the evaluated Indonesian social media dataset has been achieved.

In addition to evaluating IndoBERT, this study contributes a comparative analysis using a CNN+BiLSTM baseline trained and evaluated on the same dataset. The comparison showed that IndoBERT correctly identified a larger number of depressive-text samples and produced fewer false-negative predictions than the CNN+BiLSTM model. This finding provides additional context for interpreting the performance of the proposed approach within the scope of the dataset used in this study.

Analysis of the training and validation curves across 5 stratified cross-validation folds showed that validation loss reached its minimum value within the first few epochs and subsequently increased while training loss continued to decrease. The results suggest that overfitting began in the later epochs of training. To prevent the model to continue the training process when the performance on validation subset was no longer improving, early stopping was applied. As a result, the best checkpoints were generally obtained around the third or fourth epoch across the 5 folds.

There a re several limitations to consider in this study. The experiments relied on a single dataset collected from one social media platform, which may limit the generalizability of the findings to other datasets, social media platforms, or populations. Additionally, no resampling or class-weighting method was applied to deal with the class imbalance. This may have affected the model's ability to correctly identify the depressive-text class, as reflected in its relatively low recall in this class. The comparison was also limited to IndoBERT fine-tuning and hybrid CNN+BiLSTM model. Other Transformer-based architectures such as multilingual BERT and RoBERTa, were not evaluated. Thus, the results should be interpreted as a comparison of IndoBERT and a conventional neural baseline rather than a comprehensive comparison of a broader range of models available for this task.

Future studies could address this limitation by applying class-balancing methods and data augmentation techniques, as well as by evaluating other transformer-based architecture. These approaches could provide insight on how different modeling and imbalance data handling strategies or methods affect depression detection in Indonesian social media text.

References

[1] Katalog Data, "Katalog Data - Layanan Permintaan Data | Kementerian Kesehatan RI." Accessed: Aug. 02, 2026. [Online]. Available: https://layanandata.kemkes.go.id/katalog-data/ski/ketersediaan-data/ski-2023

[2] D. William and D. Suhartono, "Text-based Depression Detection on Social Media Posts: A Systematic Literature Review," *Procedia Computer Science*, vol. 179, pp. 582–589, Jan. 2021, doi: 10.1016/j.procs.2021.01.043.

[3] F. I. Kurniadi, N. L. P. S. P. Paramita, E. F. A. Sihotang, M. S. Anggreainy, and R. Zhang, "BERT and RoBERTa Models for Enhanced Detection of Depression in Social Media Text," *Procedia Computer Science*, vol. 245, pp. 202–209, Jan. 2024, doi: 10.1016/j.procs.2024.10.244.

[4] C. Lin *et al.*, "SenseMood: Depression Detection on Social Media," in *Proceedings of the 2020 International Conference on Multimedia Retrieval*, in ICMR '20. New York, NY, USA: Association for Computing Machinery, Jun. 2020, pp. 407–411. doi: 10.1145/3372278.3391932.

[5] C. Xin and L. Q. Zakaria, "Integrating Bert With CNN and BiLSTM for Explainable Detection of Depression in Social Media Contents," *IEEE Access*, vol. 12, pp. 161203–161212, 2024, doi: 10.1109/ACCESS.2024.3488081.

[6] N. V. Babu and E. G. M. Kanaga, "Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review," *SN COMPUT. SCI.*, vol. 3, no. 1, p. 74, Nov. 2021, doi: 10.1007/s42979-021-00958-1.

[7] S. P. Devika, M. R. Pooja, M. S. Arpitha, and R. Vinayakumar, "BERT-Based Approach for Suicide and Depression Identification," in *Proceedings of Third International Conference on Advances in Computer Engineering and Communication Systems*, A. B. Reddy, S. Nagini, V. E. Balas, and K. S. Raju, Eds., Singapore: Springer Nature, 2023, pp. 435–444. doi: 10.1007/978-981-19-9228-5_36.

[8] M. Pota, M. Ventura, H. Fujita, and M. Esposito, "Multilingual evaluation of pre-processing for BERT-based sentiment analysis of tweets," *Expert Systems with Applications*, vol. 181, p. 115119, Nov. 2021, doi: 10.1016/j.eswa.2021.115119.

[9] L. A. Widjayanto and E. B. Setiawan, "Depression Detection using Convolutional Neural Networks and Bidirectional Long Short-Term Memory with BERT variations and FastText Methods," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 3, pp. 1555–1568, Jun. 2025, doi: 10.52436/1.jutif.2025.6.3.4874.

[10] A. Raj, Z. Ali, S. Chaudhary, K. K. Bali, and A. Sharma, "Depression Detection Using BERT on Social Media Platforms," in *2024 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, Aug. 2024, pp. 228–233. doi: 10.1109/IICAIET62352.2024.10730329.

[11] I. R. Hidayat and W. Maharani, "General Depression Detection Analysis Using IndoBERT Method," *International Journal on Information and Communication Technology (IJoICT)*, vol. 8, no. 1, pp. 41–51, Aug. 2022, doi: 10.21108/ijoict.v8i1.634.

[12] A. S. Rizky and E. Y. Hidayat, "Emotion Classification in Indonesian Text Using IndoBERT," *Computer Engineering and Applications Journal*, vol. 14, no. 1, pp. 1–11, Feb. 2025, doi: 10.18495/comengapp.v14i1.494.

[13] S. S. Wibisono, M. A. Ulinuha, and S. Nur'aini, "Text Mining for Classifying Potentially Depressive Tweets on X Using IndoBERT," *J Statistika: Jurnal Ilmiah Teori dan Aplikasi Statistika*, vol. 18, no. 2, pp. 1073–1085, Dec. 2025, doi: 10.36456/jstat.vol18.no2.a10873.

[14] G. H. Martono and N. Sulistianingsih, "Fine-tuning Transformer Models for Emotion Detection in Indonesian Tweets Indicating Depression," *Vietnam J. Comp. Sci.*, pp. 1–22, Nov. 2025, doi: 10.1142/S2196888825500253.

[15] G. F. Situmorang and R. Purba, "Deteksi Potensi Depresi dari Unggahan Media Sosial X Menggunakan IndoBERT," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 2, pp. 649–661, Sep. 2024, doi: 10.47065/bits.v6i2.5496.

[16] S. Hans, "Depression and Anxiety in Twitter (ID)," Kaggle. Accessed: Feb. 01, 2026. [Online]. Available: https://www.kaggle.com/datasets/stevenhans/depression-and-anxiety-in-twitter-id

[17] A. F. Tatang and M. H. Assidiqi, "Comparative Analysis of Bidirectional Encoder Representations from Transformers Models for Twitter Sentiment Classification using Text Mining on Streamlit," *Jurnal Ilmu Komputer dan Informatika*, vol. 5, no. 2, pp. 127–142, Dec. 2025, doi: 10.54082/jiki.307.

[18] J. Huang *et al.*, "Incorporating emoji sentiment information into a pre-trained language model for Chinese and English sentiment analysis," *Intelligent Data Analysis*, vol. 28, no. 6, pp. 1601–1625, Nov. 2024, doi: 10.3233/IDA-230864.

[19] A. Bustamin, A. A. Prayogi, D. Siswanto, M. Rafrin, and A. Nurdin, "Text Normalization for Indonesian Slang Words in Sentiment Analysis Development," *ICIC Express Letters, Part B: Applications*, vol. 16, no. 2, pp. 121–129, Feb. 2025, doi: 10.24507/icicelb.16.02.121.

[20] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," in *Proceedings of the 28th International Conference on Computational Linguistics*, Barcelona, Spain (Online): International Committee on Computational Linguistics, Dec. 2020, pp. 757–770. doi: 10.18653/v1/2020.coling-main.66.

[21] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," May 24, 2019, *arXiv*: arXiv:1810.04805. doi: 10.48550/arXiv.1810.04805.

[22] B. Wilie *et al.*, "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," in *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*, K.-F. Wong, K. Knight, and H. Wu, Eds., Suzhou, China: Association for Computational Linguistics, Dec. 2020, pp. 843–857. doi: 10.18653/v1/2020.aacl-main.85.

[23] A. Vaswani *et al.*, "Attention is All you Need," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017. Accessed: Feb. 28, 2026. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[24] F. Muftie and M. Haris, "IndoBERT Based Data Augmentation for Indonesian Text Classification," in *2023 International Conference on Information Technology Research and Innovation (ICITRI)*, Aug. 2023, pp. 128–132. doi: 10.1109/ICITRI59340.2023.10250061.

[25] C. Sun, X. Qiu, Y. Xu, and X. Huang, "How to Fine-Tune BERT for Text Classification?," in *Chinese Computational Linguistics*, M. Sun, X. Huang, H. Ji, Z. Liu, and Y. Liu, Eds., Cham: Springer International Publishing, 2019, pp. 194–206. doi: 10.1007/978-3-030-32381-3_16.

[26] T. Wolf *et al.*, "Transformers: State-of-the-Art Natural Language Processing," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Q. Liu and D. Schlangen, Eds., Online: Association for Computational Linguistics, Oct. 2020, pp. 38–45. doi: 10.18653/v1/2020.emnlp-demos.6.

[27] A. Vaswani *et al.*, "Attention is All you Need," *Advances in Neural Information Processing Systems*, vol. 30, pp. 5998–6008, 2017.