Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 3, Juni 2025, Page. 1555-1568
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4874

Depression Detection using Convolutional Neural Networks and Bidirectional Long Short-Term Memory with BERT variations and FastText Methods

Leonardus Adi Widjayanto*[1], Erwin Budi Setiawan[2]

[1,2]Informatics Study Program, faculty of Informatics, Telkom University, Indonesia

Email: [1]leonardusadi@telkomuniversity.ac.id

Received : Jun 12, 2025; Revised : Jun 26, 2025; Accepted : Jun 26, 2025; Published : Jun 30, 2025

Abstract

Depression has become a significant public health concern in Indonesia, with many individuals expressing mental distress through social media platforms like Twitter. As mental health issues like depression are increasingly prevalent in the digital age, social media provides a valuable avenue for automated detection via text, though obstacles such as informal language, vagueness, and contextual complexity in social media complicate precise identification. This study aims to develop an effective depression detection model using Indonesian tweets by combining Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BiLSTM). The dataset consisted of 58,115 tweets, labeled into depressed and non-depressed categories. The data were preprocessed, followed by feature extraction using BERT and feature expansion using FastText. The FastText model was trained on three corpora: Tweet, IndoNews, and combined Tweet+IndoNews corpus; the total corpus will be 169,564 entries. The best result was achieved by BiLSTM model with 84.67% accuracy, a 1.94% increase from the baseline, and the second best was the BiLSTM-CNN hybrid model achieved 84.61 with an accuracy increase of 1.7% from the baseline. These result indicate that combining semantic feature expansion with deep learning architecture effectively improves the accuracy of depression detection on social media platforms. These insights highlight the importance of integrating semantic enrichment and contextual modeling to advance automated mental health diagnostics in Indonesian digital ecosystems.

Keywords : *BERT, BiLSTM, CNN, Depression detection, FastText, Social Media.*

1. INTRODUCTION

Depression is recognized as a medical condition, and it is one of the most prevalent mental illness affecting millions of people globally [1]. With the improvement of people's material living conditions, mental health problems have increasingly drawn widespread attention. Depression, as a major category of mood disorders. It is primary marked by a persistent low mood, along with loss of interest and pleasure [2]. Individuals with depressive tendencies are often overwhelmed by negative thoughts and show a noticeable bias towards negative stimuli. They tend to use words characterized by rejection and negative expressions, reflecting emotions such as sadness, stress, lack of motivation, or dissatisfaction [3]. These complex conditions significantly disrupt cognitive functioning, often resulting in poor concentration, difficulty in decision-making, and a pessimistic view of the future [4].

Severe consequences of depression, reflected over 700.000 suicide cases each year, highlight the critical need to understand better and identify early indicators of depressive behavior [5]. Timely

Vol. 6, No. 3, Juni 2025, Page. 1555-1568
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4874

recognition and intervention can reduce symptom severity, improve health outcomes, and extend an individual's lifespan [6][7]. It also helps minimize the negative effects on well being, work productivity, and social relationships. Social media platforms such as X have emerged ass space where individuals often express their emotional states through short, personal posts. The vast volume of publicly available content on X provides a rich resource for mental health analysis. The abundance of user-generated data enables algorithms to infer individuals emotional states from linguistic cues [8]. NLP models can detect patterns and sentiment signals indicative of depression [9]. Meta-analyses consistently show that depressed individuals language is more likely to contain cognitive distortions, such as those involving absolutist phrases and personalization, which are overrepresented compared to non-depressed user [10].

Techniques for detecting depression can effectively identify individuals who are experiencing or are at significant risk of developing depression. Currently, depression is typically diagnosed using questionnaires and consultations with mental health experts [11]. These approaches have limitations in scope, cost, and accessibility. Leveraging the widespread use of social media, depression detection can be automated, offering wider scope and more cost-efficiency. However, the biggest challenge in social media analysis is the complexity of textual data and the need to understand the context [12].

Various methods have been developed for detecting depression utilizing large-scale social media data. Consequently, accurately constructing user emotional state representations and identifying critical sentiment information from a massive volume of posted content is very important [3]. This research [13] proposes a hybrid deep-learning framework that combines Convolutional Neural Network (CNN) for feature extraction with a Bidirectional Long Short Term Memory (BiLSTM) layer capturing sequential dependencies in tweet text. Tweets were first converted into TF-IDF feature vectors, which were then processed through CNN filters (kernel sizes 3–5) to detect important n-gram patterns. The CNN output was passed into the BiLSTM layer, and the final classification was performed using a softmax layer. This model, tested on an English-language depression tweet dataset, achieved 94.28% accuracy, outperforming the CNN only baseline of 91.73%.

FastText, a subword embedding model effective for informal language [14][15], has been used to enhance CNN-BiLSTM architectures through vocabulary expansion [16]. It was used to enhance CNN-BiLSTM architectures by expanding vocabulary coverage, especially for slang and misspelled words in tweets. The enriched feature vectors were then fed into a CNN-BiLSTM model, which combines CNN's ability to detect local patterns and BiLSTM's capacity for modeling long-term dependencies. The CNN-BiLSTM achieved 80.55% accuracy, while a BiLSTM-CNN variant scored 80.35%. These results represented performance boosts of 1.86% and 2.90% over the baseline, respectively.

Bidirectional Encoder Representations from Transformers (BERT) is a pre-trained transformer model known for producing deep, context-aware language representations. It is particularly effective at capturing nuanced meanings in short texts, such as tweets [17]. BERT utilizes a transformer-based architecture, excelling in understanding intricate contextual relationships within text data and transforming semantically similar words into vector representations with similar distances [18]. This research [19], fine-tuned several transformer models—BERT, RoBERTa, BERTweet, and MentalBERT on depression-related datasets from Reddit and Twitter. After preprocessing and tokenization, the models were trained using standard hyperparameters adapted to mental health contexts. RoBERTa achieved the highest accuracy individually, followed by BERTweet and BERT. An ensemble model "GT", which averaged the predictions of RoBERTa and BERTweet, further improved performance to 87.3%, surpassing the best single model (RoBERTa at 86.6%) by 0.7%.

Building on existing research, this study combines the strengths of previous methods into a unified framework. Therefore, this study merges BERT for feature extraction, FastText for feature Expansion, and hybrid method CNN-BiLSTM architecture to enhance semantic understanding, accuracy, and model robustness in detecting depression in Indonesian languages from social media posts. This approach

Vol. 6, No. 3, Juni 2025, Page. 1555-1568
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4874

enhances the model's semantic understanding and classification performance. The findings are expected to support early mental health detection, assist healthcare professionals in identifying signs of depression on social media, and contribute to mental health awareness efforts in Indonesia.

2. METHOD

The flowchart shown in Figure 1 illustrates the methodology of depression detection on social media, which consists of data crawling, labeling, preprocessing, and data splitting. Then we continue the process with feature extraction using BERT and feature expansion using FastText.

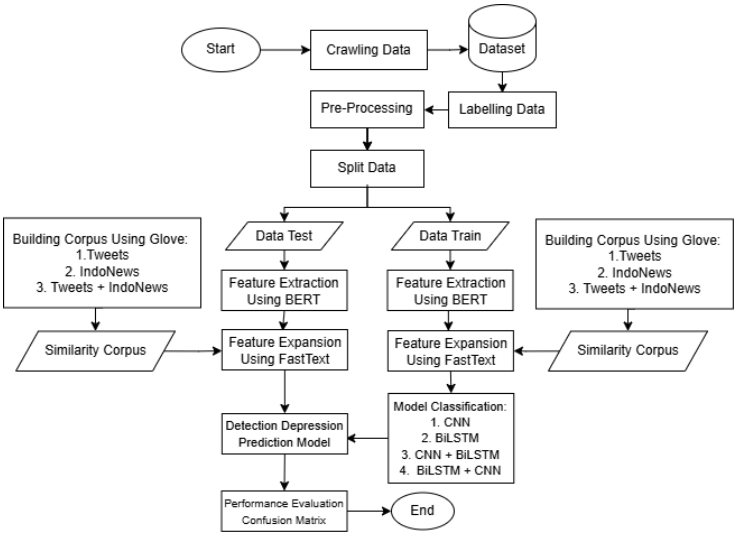


Figure 1 System workflow using a hybrid CNN-BiLSTM architecture

2.1. Data Crawling from Twitter

The process of collecting data or information from websites is called data crawling [20]. Data was collected from Twitter using the platform's API, focusing on posts in the Indonesian language that contain keywords and hashtags commonly associated with depression (e.g., #depresi, #sedih, #capek). Only publicly available tweets were considered, adhering to ethical guidelines and privacy policies. As a result, a total of 58,115 tweet entries were collected. The distribution of keywords used during this process is shown in Table I.

Table I QUANTITY OF CRAWLED DATASET

Keywords	Quantity
cemas	9,245
sendirian	6,842
capek	7,189
lelah	6,318
depresi	7,915
sedih	8,427
frustasi	6,031
ngerasa gagal	6,148
Total	**58,115**

Vol. 6, No. 3, Juni 2025, Page. 1555-1568
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4874

2.2. Data Labeling Based on Depression Indicators

Data labeling is the process of assigning categories or tags to unprocessed data after data crawling is completed. In this study, the collected tweet data was labeled into two classes: 1 for depressed and 0 for non-depressed, as presented in Table II. To maintain objectivity and precision, the labeling process is executed by three annotators. Engaging several annotators ensures a variety of viewpoints, reducing bias or errors that might occur if only one person is involved. Each tweet was carefully assessed according to its contextual meaning to ensure that the chosen label correctly captures the psychological expression conveyed.

Table II A MOUNT OF DATA CRAWLED

Category	Label	Quantity
Depressed	1	29,421
Non-Depressed	0	28,694
Total		58,115

2.3. Data Pre-Processing

After the crawling process is completed, the next step is preprocessing to convert unstructured data into structured data [16]. The tweet data that has been collected through the crawling process often contains noise [21], which is data that does not have relevant information for analysis purposes. This process consists of several steps:

1) Data Cleansing: removing unnecessary elements such as mentions, hashtags, URLs, numbers, symbols, and special characters using regular expressions.
2) Case Folding: converting letters to lowercase to maintain consistency and avoid discrepancies caused by differences in uppercase and lowercase letters.
3) Normalization: correcting spelling errors and converting words to their original form.
4) Stopwords removal: performed using the Sastrawi library in Python to remove irrelevant common words, ensuring the meaning of the words remains relevant.
5) Stemming: also using the Sastrawi library to convert words to their base form by removing affixes.
6) Tokenization: breaking sentences into individual words using the NLTK library to enable further analysis.

2.4. Dataset Splitting

Data splitting is divided into two parts: training data for training the hybrid CNN-BiLSTM model and testing data for evaluating the model's performance.

2.5. Contextual Feature Extraction using BERT

Bidirectional Encoder Representations from Transformers (BERT) was utilized to capture contextual embeddings of the text data. BERT employs a tokenizer that converts input text into embeddings, which are then processed through multiple transformer layers to generate contextualized representations of words[22]. Transformer networks, exemplified by BERT, have revolutionized natural language processing by enabling machines to discern intricate linguistic patterns, and these models are adept at capturing subtle semantic relationships within text [23]. BERT's ability to understand context in both directions enhances the semantic understanding of the text [24], which is vital for detecting nuanced expressions of depression.

2.6. Feature Expansion using FastText

Vol. 6, No. 3, Juni 2025, Page. 1555-1568
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4874

FastText, developed by Facebook AI Research, augments traditional word embeddings by incorporating sub word information, thereby facilitating the handling of out of vocabulary words and capturing morphological nuances. This attribute is particularly advantageous for the analysis of social media text, which often contains slang, misspellings, and creative language use [16][17]. In this study, FastText is used for feature expansion by calculating word similarity within each corpus, aiming to reduce vocabulary mismatches commonly found in informal texts like tweets. The methodology begins by collecting raw data from social media platform X and IndoNews articles, followed by extensive preprocessing to normalize, clean, and structure the data for embedding training. The summary of the corpora used for FastText training is provided in Table III. Following the training phase, FastText evaluates word similarity to identify semantically related terms.

Table III QUANTITY OF CORPUS

Corpus Similarity	Quantity
Tweet	58,115
IndoNews	111,449
Tweet+IndoNews	169,564

2.7. Classification Model

2.7.1. Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNNs) have been effectively applied to text classification tasks due to their ability to capture local patterns and hierarchical features within textual data. The convolutional layers can automatically extract relevant features from the raw input data during network training and use them to classify inputs [25]. By applying convolutional filters over word embeddings, CNN can detect n-gram features, enabling the model to recognize meaningful phrases and contextual information [26]. As illustrated in Figure 2, this study, the model employs three 1D convolutional layers with kernel sizes of 3, 5, and 7, each using the ReLU function to introduce non-linearity. Following each convolutional operation, MaxPooling layers are applied to reduce the dimensionality of feature maps, while a dropout layer with a 0.5 rate is incorporated to prevent overfitting. The feature maps are flattened and forwarded to the next processing phase. The model is trained using a learning rate of 0.001 and a batch size of 64, chosen to optimize both the speed of training and the stability of convergence.

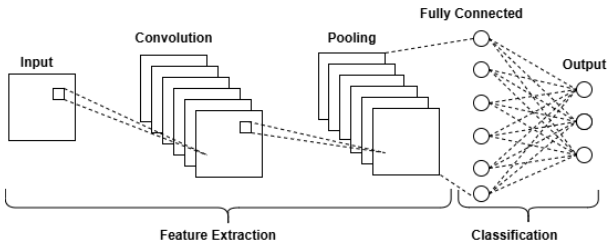


Figure 2 CNN model architecture

2.7.2. Bidirectional Long Short-Term Memory (BiLSTM)

Bidirectional Long Short-Term Memory (BiLSTM) networks are advanced recurrent neural network architectures that process sequences in both forward and backward directions, effectively capturing contextual information from both past and future states, thus enabling a more comprehensive understanding of the input data [17]. This bidirectional processing enables BiLSTMs to model complex

Vol. 6, No. 3, Juni 2025, Page. 1555-1568
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4874

dependencies within text, making them particularly effective for tasks such as sentiment analysis and topic classification. By assigning different weights to different words in a sentence, attention mechanisms enable the model to prioritize informative words, thereby improving classification accuracy. BiLSTMs are adept at learning temporal features from sequential data using memory cells and gating mechanisms [20][21]. Figure 4 presents the BiLSTM architecture applied in this study, showing process sequences in both forward and backward directions. The feature-extracted input vector is reformatted into a time-distributed structure and introduced into a Bidirectional LSTM layer comprising 64 units. Subsequently, the output from the BiLSTM layer is directed to a dense layer with 64 neurons employing the ReLU activation function. To address potential overfitting, dropout layers with a rate of 0.5 are applied following both the recurrent and dense layers. The final output layer utilizes a sigmoid activation function to facilitate binary classification.

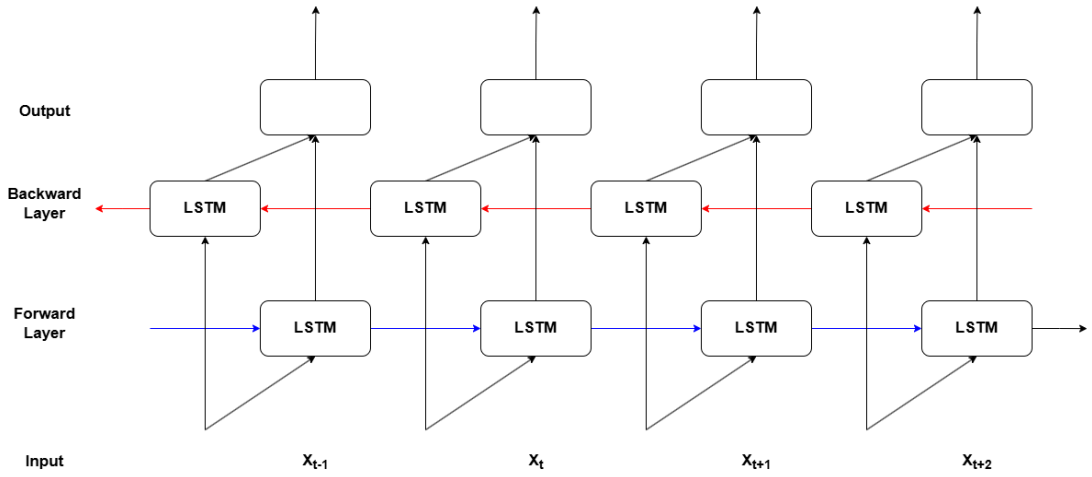


Figure 3 BiLSTM model architecture

2.7.3. Hybrid Model

Hybrid models combine the local feature detection strengths of Convolutional Neural Networks (CNNs) with the sequence-modeling capabilities of Bidirectional Long Short-Term Memory (BiLSTM) networks to capture both spatial and temporal patterns in text data [29]. In such architectures, convolutional layers first extract n-gram–level features via multiple filter sizes, while subsequent MaxPooling layers reduce dimensionality and highlight the most salient local cues. The pooled feature maps are then fed into BiLSTM layers, processing the sequence forwards and backwards to model long-range dependencies and contextual semantics across the entire input sequence [22][23]. The CNN model excels at identifying local features, while the BiLSTM captures the sequential flow of information, preserving both past and future dependencies for each token, making it highly suitable for modeling nuanced context. By Integrating CNN and BiLSTM in a parallel configuration, they can provide valuable information for each other [32]. There are two main configurations for combining CNN and BiLSTM layers: CNN-BiLSTM and BiLSTM-CNN. The hybrid model CNN-BiLSTM illustrated in Figure 4, where convolution occurs first followed by sequences modeling. The hybrid architecture leverages complementary strengths to improve classification performance. Both architectures are implemented and evaluated in this study to explore which sequence yields better performance for the depression detection task.

Vol. 6, No. 3, Juni 2025, Page. 1555-1568
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4874

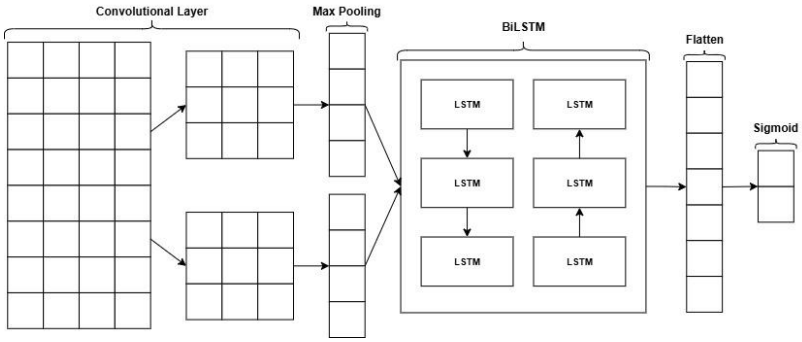


Figure 4 Hybrid CNN-BiLSTM model architecture

3. RESULT

This study conducted a series of experiments to evaluate model performance and identify the optimal configuration for depression detection. This section presents

1. Scenario 1 tests the baseline model to identify the optimal split ratio.
2. Scenario 2 identifies the optimal layers for BERT, based on results from the baseline.
3. Scenario 3 examines the effect of different n-gram combinations on feature extraction and model performance.
4. Scenario 4 implements FastText feature expansion using three corpus types and evaluates top-1, top-5, and top-10 similarity terms.

3.1. Scenario 1

In the first scenario, data was divided using ratios of 70:30, 80:20, and 90:10 for training and testing. The main objective is to evaluate the model's performance on the "baseline" assessment, determine the most effective data split ratio, and maximize accuracy. As shown in Table IV, the CNN model reached its highest accuracy of 82.63% with an 80:20 split, whereas the BiLSTM model outperformed it with 82.73% at a 90:10 split. The top overall accuracy of 82.91% was recorded by the BiLSTM-CNN model using the 90:10 split.

Table IV SCENARIO 1: SPLIT RATIO

Split Ratio	Accuracy (%)			
	CNN	**BiLSTM**	**CNN-BiLSTM**	**BiLSTM-CNN**
70:30	81.05	81.38	81.59	81.85
80:20	**82.63**	81.79	82.36	82.22
90:10	82.03	**82.73**	**82.82**	**82.91**

In summary, the optimal split ratios for baseline testing were determined, 80:20 for CNN, 90:10 for BiLSTM, BiLSTM-CNN, and CNN-BiLSTM. These ratios will be used in subsequent scenarios to ensure consistency and optimal conditions.

3.2. Scenario 2

This scenario evaluates the impact of four BERT variants: TinyBERT (4 layers) for lightweight processing in limited-resource environments [33], DistilBERT (6 layers) for efficient performance with reduced complexity [34], BERT-Base (12 layers) as the standard model for capturing rich contextual patterns, and BERT-Large (24 layers) for maximum modeling capacity and deep contextual understanding [35]. These variants offer a balance between computational efficiency and representational power depending on task requirements. The evaluation results for each model paired

Vol. 6, No. 3, Juni 2025, Page. 1555-1568
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4874

with different BERT version are presented in Table V. With the 24-layer BERT-Large configuration, accuracies reached their peak: CNN at 83.05%, BiLSTM at 83.43%, CNN–BiLSTM at 83.34%, and BiLSTM–CNN at 83.86%. This indicates that the deeper BERT variant significantly enhances the model's ability to capture complex contextual patterns.

Table V SCENARIO 2: LAYERS

Layer	Accuracy (%)			
	CNN	**BiLSTM**	**CNN-BiLSTM**	**BiLSTM-CNN**
4	82.63	82.73	82.82	82.91
6	82.71	82.98	83.01	83.27
12	82.92	83.19	83.25	83.52
24	**83.05**	**83.43**	**83.34**	**83.86**

To conclude, the BiLSTM-CNN model, when combined with BERT-Large, consistently outperforms other configurations, establishing the most effective pairing for deep contextual understanding.

3.3. Scenario 3

In the third scenario, testing the different combinations of n-grams affects the text categorization model. Tested combinations include: unigram, bigram, trigram, unigram+bigram, bigram+trigram, and allgram. The best result from split ratio and layer will be used in this test. Accuracy results are summarized in Table VI. Showcasing the accuracy percentages for the CNN, BiLSTM, CNN-BiLSTM, and BiLSTM-CNN models across different N-gram configurations. The CNN model achieved its highest accuracy of 83.05% with the unigram approach, while the BiLSTM model surpassed this with 83.45% using the same allgram setup. The top overall accuracy of 83.86% was recorded by the BiLSTM-CNN model with Unigram. Meanwhile, the CNN-BiLSTM model attained its best result with unigram at 83.34%, though it also showed solid performance with allgram at 82.58%.

Table VI SCENARIO 3: N-GRAM

N-Gram	Accuracy (%)			
	CNN	BiLSTM	CNN-BiLSTM	BiLSTM-CNN
Unigram	**83.05**	83.43	**83.34**	**83.86**
Bigram	78.92	80.51	79.15	81.63
Trigram	72.34	73.98	74.67	76.21
Uni-Bigram	81.73	82.19	80.93	83.08
Bi-Trigram	77.45	78.12	75.82	79.56
Allgram	82.45	**83.45**	82.58	83.53

Allgram provides a balanced performance slightly below unigram. These N-gram variations will guide the baseline for further analyses. In summary, unigram features yielded the best results across most models, particularly benefiting BiLSTM–CNN, while allgram configurations offered balanced but slightly lower accuracy.

3.4. Scenario 4

In the fourth scenario, FastText feature expansion is applied using similarity corpus constructed from three distinct data types: tweets, Indonews, and a combination of tweets and Indonews. Each data type undergoes three evaluations, specifically at Top 1, Top 5, and Top 15 levels. The outcomes of this scenario are detailed in Table VII, providing a comprehensive overview of the performance across these configurations. All models in scenario 4 experienced an increase in accuracy, especially the BiLSTM model which increased by 1.22% on the corpus built with the tweet and Top 1 datasets. The CNN model

Vol. 6, No. 3, Juni 2025, Page. 1555-1568
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4874

increased by 0.54% on the tweet+news and Top 1 datasets, CNN-BiLSTM increased by 0.47% on the tweet+news and Top 1 datasets, and BiLSTM-CNN increased by 0.47% on the news and Top 5 datasets.

Table VII SCENARIO 4: FASTTEXT FEATURE EXPANSION

Model	Rank	Accuracy (%)		
		Tweet	IndoNews	Tweet+IndoNews
CNN	Top 1	82.74	82.74	**83.59**
	Top 5	82.52	83.56	82.76
	Top 10	79.57	82.77	82.19
BiLSTM	Top 1	**84.67**	83.66	83.82
	Top 5	82.52	83.08	83.15
	Top 10	83.31	82.35	82.58
CNN-BiLSTM	Top 1	82.89	82.32	**83.81**
	Top 5	81.17	81.79	83.34
	Top 10	79.58	82.02	82.81
BiLSTM-CNN	Top 1	83.65	84.12	84.37
	Top 5	83.08	**84.61**	84.61
	Top 10	81.36	83.76	83.89

In conclusion, FastText expansion provides consistent gains across all models, with the Tweet-based corpus and smaller top-k settings (Top 1 or Top 5) yielding the most noticeable improvements, especially for BiLSTM-based models.

4. DISCUSSION

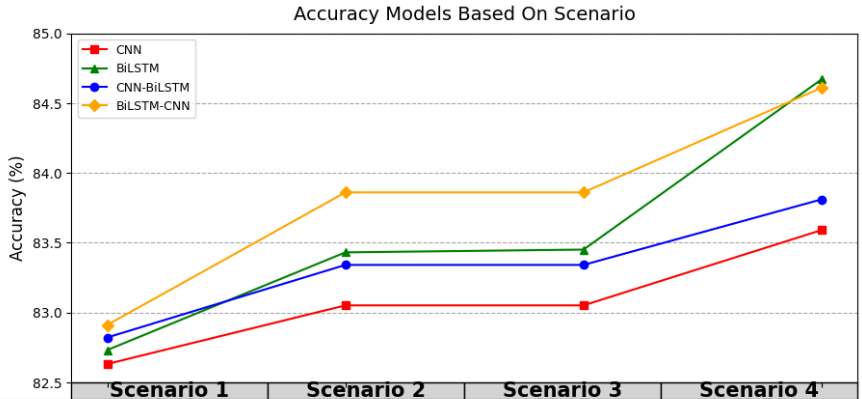
This study carried out multiple experimental scenarios to evaluate the effectiveness of different configurations in detecting depression from Indonesian language tweets. Statistical significance tests, including Z-values and P-values, were employed to determine whether the performance differences between scenarios were meaningful. As shown in Table VIII, there were no statistically significant improvements between Scenario 2 and Scenario 3 ($Z = 1.0$, $P = 0.3910$), indicating that the changes applied in Scenario 3 did not produce a meaningful effect on model accuracy. However, significant gains were observed in transitions from Scenario 1 to Scenario 2 ($Z = 5.5680$, $P = 0.0114$), Scenario 3 to Scenario 4 ($Z = 4.4046$, $P = 0.0217$), and especially from Scenario 1 to Scenario 4 ($Z = 5.6152$, $P = 0.0111$). These results suggest that each improvement in those transitions led to statistically reliable accuracy gains. Notably, the transition from Scenario 1 to Scenario 4 shows the most substantial and consistent performance improvement, underscoring the effectiveness of semantic feature expansion and combined model architectures in enhancing depression detection accuracy. Overall, the significance test results validate the experimental setup and confirm that the enhancements introduced in Scenario 4 yield a considerable and statistically supported impact on model performance. This provides strong evidence for the superiority of the final configuration in this study.

Table VIII SIGNIFICANCE TEST IN SCENARIOS

Scenario Transitions	Z-Values	P-Values	Significant?
S1 → S2	5.5680	0.0114	TRUE
S2 → S3	1.0	0.3910	FALSE
S3 → S4	4.4046	0.0217	TRUE
S1 → S4	5.6152	0.0111	TRUE

Vol. 6, No. 3, Juni 2025, Page. 1555-1568
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4874

In this study, a series of test scenarios were conducted to determine the optimal model configuration. As illustrated in the Figure 5, the accuracy comparison of the CNN, BiLSTM, CNN-BiLSTM, and BiLSTM-CNN models across four scenarios is presented. Scenario 1 served as the baseline, with initial accuracies of 82.63% for CNN, 82.73% for BiLSTM, 82.82% for CNN-BiLSTM, and 82.91% for BiLSTM-CNN. In Scenario 2, variations of BERT features resulting an improved accuracy across all models. CNN improved to 83.05%, BiLSTM to 83.43%, CNN-BiLSTM to 83.34%, and BiLSTM-CNN to 83.86%. This shows that adding more features positively impacts model performance. Scenario 3, which explored various n-gram combinations, did not significantly affect accuracy, with most models maintaining similar performance to Scenario 2. CNN remained at 83.05%, BiLSTM at 83.45%, CNN-BiLSTM at 83.34%, and BiLSTM-CNN at 83.86%. In Scenario 4, semantic feature expansion using FastText led to the highest accuracy improvements across all models. CNN reached 83.59%, BiLSTM improved to 84.67%, CNN-BiLSTM achieved 83.81%, and BiLSTM-CNN recorded the highest accuracy at 84.61%. These results indicate that the integration of semantic feature expansion significantly boosts model performance, with the BiLSTM-CNN model consistently outperforming the others in the final scenario.



	'Scenario 1	Scenario 2	Scenario 3	Scenario 4'
CNN	82.63	83.05	83.05	83.59
BiLSTM	82.73	83.43	83.45	84.67
CNN-BiLSTM	82.82	83.34	83.34	83.81
BiLSTM-CNN	82.91	83.86	83.86	84.61

Figure 5 Accuracy between scenarios

This research [16], proposes a hybrid CNN-BiLSTM model enhanced with TF-IDF feature extraction and FastText feature expansion. FastText embeddings are trained on a combined Tweet + IndoNews corpus to enrich vocabulary coverage, particularly capturing informal slang and misspellings. The CNN layers then extract local n-gram patterns, while BiLSTM units model long-range dependencies in the enriched feature vectors. This integration yields an accuracy of 80.55% for CNN–BiLSTM and 80.35% for BiLSTM–CNN, marking improvements of 1.86% and 2.90% over their respective CNN and BiLSTM baselines. However, these subword embeddings alone may miss broader contextual cues essential for depression detection; therefore, our work augments this approach with BERT-based contextual embeddings that capture both syntactic and semantic relationships across sentences. Out works, combined with FastText morphological robustness, raises BiLSTM accuracy to 84.67%—an absolute gain of 4.12% demonstrating the value of fusing subword and deep contextual features to improve model generalization on informal social media text.

This research [13], Kour and Gupta proposed a hybrid framework that using TF-IDF for feature extraction, follower by CNN layers and BiLSTM model. Their model achieved 94.28% accuracy outperforming single CNN model with 91.73% accuracy. While our works utilizing BERT for feature extraction, FastText for feature expansion, and CNN-BiLSTM model achieved 84.61% on BiLSTM-CNN and 84.67% on BiLSTM accuracy on Indonesian tweets. While the result is behind the Kour and Gupta research, our model addresses unique challenges in the Indonesian language, such as limited

Vol. 6, No. 3, Juni 2025, Page. 1555-1568
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4874

pretrained resources, highly informal syntax, slang, and spelling variations. Thus, although absolute accuracy is lower, our findings mark a significant achievement in a low-resource language context, demonstrating the hybrid model's versatility and the effectiveness of contextual and subword-level enrichment.

In other research [19], Tavchioski and colleagues fine-tuned multiple transformer variants using BERT, RoBERTa, BERTweet, and mentalBERT on Reddit and Twitter dataset for detection depression. They found RoBERTa achieved the highest single-model accuracy at 86.6%, and an ensemble of RoBERTa + BERTweet further improved accuracy to 87.3%, a 0.7% gain over the strongest individual model. While our best model is BiLSTM with BERT + FastText reaches 84.67%, it trails the ensemble by 2.6%. This gap underscores the performance advantage of transformer ensembles, our research focuses on the underexplored domain of Indonesian language social media rather than English content, which presents unique technical and linguistic challenges.

Importantly, the methods and findings presented here are not limited to Twitter or depression detection alone. The same of all combined methods could be applied to other social platforms, such as Instagram captions or Tiktok comments, where text fragments and slang are used. Moreover, by retraining with annotated data for related mental health concerns (e.g., anxiety, suicidal thoughts), this framework can generalize to detect a broader spectrum of psychological conditions. Such adaptability suggests strong potential for multi domain mental health monitoring across diverse online environments.

The findings of this study demonstrate the potential for deploying automated depression detection models as part of mental health surveillance systems on social media platforms. Enabling early identification of individuals exhibiting signs of psychological distress and timely interventions, particularly valuable in low resource contexts. However, their real world deployment demands careful ethical consideration, as risks related to privacy, consent, algorithmic bias, and user stigmatization may undermine their utility. Ensuring responsible use requires transparent governance, human oversight, and safeguards that uphold user autonomy, positioning these systems not as replacements but as ethically grounded augmentations to professional mental health care.

5. CONCLUSION

This research investigated the performance of hybrid CNN-BiLSTM models in detecting depression from Indonesian language tweets supported by BERT-based features and FastText feature expansion. The dataset consisted of 58,115 tweets, collected using eight mental health-related keywords, and evenly labeled into depressed (29,421) and non-depressed (28,694) categories. The dataset was expanded with FastText embeddings for Tweets, IndoNews, and their combination, totaling 169,564 instances. Four experimental scenarios were used to determine the optimal configuration. Starting with variations in train-test split ratios, variations of BERT layers, n-gram ranges, and semantic enrichment with FastText. This study offers a valuable contribution to the fields of informatics and computer science, especially in the areas of mental health and social media analysis, by utilizing Indonesian-language data and integrating BERT with FastText. Among the tested models, BiLSTM model achieved the highest accuracy of 84.67%, improving by 1.94% over the baseline, followed closely by the BiLSTM-CNN hybrid with 84.61% accuracy and a 1.7% improvement. This approach enhances both the accessibility and efficiency of mental health interventions. Future work may explore deeper contextual embeddings or domain adaptation to further optimize classification outcomes. This study contribution lies in advancing NLP applications for Indonesia languages in mental health contexts, with potential integration into online psychological support platforms to assist early intervention and digital screening efforts.

CONFLICT OF INTEREST

The authors declares that there is no conflict of interest between the authors or with research object in this paper.

REFERENCES

[1] Vandana, N. Marriwala, and D. Chaudhary, "A hybrid model for depression detection using deep learning," *Meas. Sensors*, vol. 25, no. December 2022, p. 100587, 2023, doi: 10.1016/j.measen.2022.100587.

[2] Z. Wang, L. Chen, L. Wang, and G. Diao, "Recognition of Audio Depression Based on Convolutional Neural Network and Generative Antagonism Network Model," *IEEE Access*, vol. 8, pp. 101181–101191, 2020, doi: 10.1109/ACCESS.2020.2998532.

[3] G. Rao, Y. Zhang, L. Zhang, Q. Cong, and Z. Feng, "MGL-CNN: A Hierarchical Posts Representations Model for Identifying Depressed Individuals in Online Forums," *IEEE Access*, vol. 8, pp. 32395–32403, 2020, doi: 10.1109/ACCESS.2020.2973737.

[4] X. Zhang *et al.*, "The influence of genetic and acquired factors on the vulnerability to develop depression: a review," *Biosci. Rep.*, vol. 43, no. 5, pp. 1–17, 2023, doi: 10.1042/BSR20222644.

[5] World Health Organization, "Depressive disorder (depression)," Depressive disorder (depression). Accessed: Apr. 29, 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression

[6] C. G. Davey and P. D. Mcgorry, "Early intervention for depression in young people : a blind spot in mental health care," *The Lancet Psychiatry*, vol. 0366, no. 18, 2018, doi: 10.1016/S2215-0366(18)30292-X.

[7] Y. Chen *et al.*, "Effectiveness of health checkup with depression screening on depression treatment and outcomes in middle-aged and older adults : a target trial emulation study," *Lancet Reg. Heal. - West. Pacific*, vol. 43, p. 100978, 2024, doi: 10.1016/j.lanwpc.2023.100978.

[8] A. A. Jamali, C. Berger, and R. J. Spiteri, "Momentary Depressive Feeling Detection Using X (Formerly Twitter) Data: Contextual Language Approach," *Jmir Ai*, vol. 2, p. e49531, 2023, doi: 10.2196/49531.

[9] B. G. Bokolo and Q. Liu, "Deep Learning-Based Depression Detection from Social Media: Comparative Evaluation of ML and Transformer Techniques," *Electron.*, vol. 12, no. 21, 2023, doi: 10.3390/electronics12214396.

[10] K. C. Bathina, L. Lorenzo-luaces, L. A. Rutter, B. Complexity, and B. Sciences, "Depressed individuals express more distorted thinking on social media," no. Cd, pp. 1–21, 2020, doi: https://doi.org/10.48550/arXiv.2002.02800.

[11] A. Ranjith Kumar, K. Aditya, S. Antony Joseph Raj, and V. Nandhakumar, "Depression Detection Using Optical Characteristic Recognition and Natural Language Processing in SNS," *4th Int. Conf. Comput. Commun. Signal Process. ICCCSP 2020*, 2020, doi: 10.1109/ICCCSP49186.2020.9315254.

[12] D. Liu, X. L. Feng, F. Ahmed, M. Shahid, and J. Guo, "Detecting and Measuring Depression on Social Media Using a Machine Learning Approach: Systematic Review," *JMIR Ment. Heal.*, vol. 9, no. 3, pp. 1–18, 2022, doi: 10.2196/27244.

[13] H. Kour and M. K. Gupta, "An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM," *Multimed. Tools Appl.*, vol. 81, no. 17, 2022, doi: 10.1007/s11042-022-12648-y.

[14] J. Choi and S. W. Lee, "Improving FastText with inverse document frequency of subwords," *Pattern Recognit. Lett.*, vol. 133, pp. 165–172, 2020, doi: 10.1016/j.patrec.2020.03.003.

[15] N. Sabharwal and A. Agrawal, "Introduction to Word Embeddings," in *Hands-on Question Answering Systems with BERT*, Apress, 2021, p. 41. doi: 10.1007/978-1-4842-6664-9_3.

[16] M. A. S. Nasution and E. B. Setiawan, "Enhancing Cyberbullying Detection on Indonesian Twitter: Leveraging FastText for Feature Expansion and Hybrid Approach Applying CNN and BiLSTM," *Rev. d'Intelligence Artif.*, vol. 37, no. 4, pp. 929–936, 2023, doi: 10.18280/ria.370413.

[17] F. Ullah, A. Alsirhani, M. M. Alshahrani, A. Alomari, H. Naeem, and S. A. Shah, "Explainable Malware Detection System Using Transformers-Based Transfer Learning and Multi-Model Visual Representation," *Sensors*, vol. 22, no. 18, 2022, doi: 10.3390/s22186766.

[18] S. F. N. Azizah, H. D. Cahyono, S. W. Sihwi, and W. Widiarto, "Performance Analysis of Transformer Based Models (BERT, ALBERT, and RoBERTa) in Fake News Detection," *2023 6th Int. Conf. Inf. Commun. Technol. ICOIACT 2023*, pp. 425–430, 2023, doi:

Vol. 6, No. 3, Juni 2025, Page. 1555-1568
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4874

10.1109/ICOIACT59844.2023.10455849.
[19] I. Tavchioski, M. Robnik-Šikonja, and S. Pollak, "Detection of depression on social networks using transformers and ensembles," 2023, doi: 10.14746/amup.9788323241775.
[20] B. A. Putri and E. B. Setiawan, "Topic Classification Using the Long Short-Term Memory (LSTM) Method with FastText Feature Expansion on Twitter," *2023 Int. Conf. Data Sci. Its Appl. ICoDSA 2023*, no. April, pp. 18–23, 2023, doi: 10.1109/ICoDSA58501.2023.10277033.
[21] V. Çetin and O. Yıldız, "A comprehensive review on data preprocessing techniques in data analysis," *Pamukkale Univ. J. Eng. Sci.*, vol. 28, no. 2, pp. 299–312, 2022, doi: 10.5505/pajes.2021.62687.
[22] R. Menon, S. Gide, S. Ghatte, S. Mendon, M. Nashipudimath, and U. G. Student, "Depression Prediction using BERT and SVM," *Int. Res. J. Eng. Technol.*, pp. 2013–2016, 2021, [Online]. Available: www.irjet.net
[23] D. Cortiz, "Exploring Transformers in Emotion Recognition: a comparison of BERT, DistillBERT, RoBERTa, XLNet and ELECTRA," pp. 1–7, 2021, [Online]. Available: http://arxiv.org/abs/2104.02041
[24] R. K. Kaliyar, A. Goswami, and P. Narang, "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach," *Multimed. Tools Appl.*, vol. 80, no. 8, pp. 11765–11788, 2021, doi: 10.1007/s11042-020-10183-2.
[25] K. M. Ang *et al.*, "Optimizing Image Classification: Automated Deep Learning Architecture Crafting with Network and Learning Hyperparameter Tuning," *Biomimetics*, vol. 8, no. 7, 2023, doi: 10.3390/biomimetics8070525.
[26] S. Soni, S. S. Chouhan, and S. S. Rathore, "TextConvoNet: a convolutional neural network based architecture for text classification," *Appl. Intell.*, vol. 53, no. 11, pp. 14249–14268, 2023, doi: 10.1007/s10489-022-04221-9.
[27] P. Nevavuori, N. Narra, P. Linna, and T. Lipping, "Crop yield prediction using multitemporal UAV data and spatio-temporal deep learning models," *Remote Sens.*, vol. 12, no. 23, pp. 1–18, 2020, doi: 10.3390/rs12234000.
[28] J. Liu, Y. Zhao, Y. Feng, Y. Hu, and X. Ma, "SeMalBERT: Semantic-based malware detection with bidirectional encoder representations from transformers," *J. Inf. Secur. Appl.*, vol. 80, p. 103690, Feb. 2024, doi: 10.1016/J.JISA.2023.103690.
[29] J. Forry Kusuma and A. Chowanda, "Indonesian Hate Speech Detection Using IndoBERTweet and BiLSTM on Twitter," *Int. J. Informatics Vis.*, vol. 7, no. September, pp. 773–780, 2023, [Online]. Available: www.joiv.org/index.php/joiv
[30] T. H. H. Aldhyani, S. N. Alsubari, A. S. Alshebami, H. Alkahtani, and Z. A. T. Ahmed, "Detecting and Analyzing Suicidal Ideation on Social Media Using Deep Learning and Machine Learning Models," *Int. J. Environ. Res. Public Health*, vol. 19, no. 19, 2022, doi: 10.3390/ijerph191912635.
[31] A. Abdurrahim and D. H. F. Fudholi, "Mental Health Prediction Model on Social Media Data Using CNN-BiLSTM," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 4, no. 1, 2024, doi: 10.22219/kinetik.v9i1.1849.
[32] W. Li, L. Zhu, Y. Shi, K. Guo, and E. Cambria, "User reviews: Sentiment analysis using lexicon integrated two-channel CNN–LSTM family models," 2020. doi: 10.1016/j.asoc.2020.106435.
[33] X. Jiao, "TinyBERT : Distilling BERT for Natural Language Understanding," pp. 4163–4174, 2020.
[34] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT , a distilled version of BERT : smaller , faster , cheaper and lighter," pp. 2–6, 2019.
[35] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, Association for Computational Linguistics (ACL), 2019, pp. 4171–4186.

Vol. 6, No. 3, Juni 2025, Page. 1555-1568
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4874