BAREKENG: Journal of Mathematics and Its Applications
June 2026 Volume 20 Issue 2 Page 1719–1730
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss2pp1719-1730

HYBRID INTEGRATION OF BERT AND BILSTM MODELS FOR SENTIMENT ANALYSIS

Nicolas Ray Amarco Tambunan[1*], **Dewi Retno Sari Saputro**[2], **Purnami Widyaningsih**[3]

[1,2,3]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Sebelas Maret*
Jln. Ir. Sutami No. 36 Kentingan, Surakarta, 57126, Indonesia

*Corresponding author's e-mail: * **nicolasr4y@student.uns.ac.id***

Article Info

Article History:

Received: 20th August 2025
Revised: 5th September 2025
Accepted: 6th October 2025
Available Online: 26th January 2026

Keywords:

BERT;
BiLSTM;
NLP;
RNN;
Sentiment analysis.

ABSTRACT

The rapid growth of sentiment analysis research has driven increasing interest in deep learning models, particularly transformer-based architectures such as BERT and recurrent neural networks like BiLSTM. While both approaches have shown substantial success in text classification tasks, each presents distinct strengths and limitations. This study aims to analyze the integration of BERT and BiLSTM models to enhance sentiment classification performance by combining contextual and sequential learning. A bibliometric analysis was conducted using VosViewer based on Scopus-indexed publications from 2020 to 2025, identifying four major thematic clusters related to transformer modeling, recurrent architectures, hybrid integration, and methodological advancements. Comparative findings from benchmark datasets, including SST-2, IMDb, and Yelp Reviews, indicate that hybrid BERT–BiLSTM models achieve superior accuracy compared to single models, reaching up to 97.67% on the IMDb dataset. However, this improvement is associated with increased computational complexity. The proposed framework reinforces the integration between BERT's contextual embeddings and BiLSTM's sequential modeling, offering a foundation for developing adaptive, and multilingual sentiment analysis systems. The results highlight future directions in optimizing hybrid architectures for efficiency, cross-lingual adaptability, and domain-specific sentiment understanding.

How to cite this article:

N. R. A. Tambunan, D. R. S. Saputro and P. Widyaningsih., "HYBRID INTEGRATION OF BERT AND BILSTM MODELS FOR SENTIMENT ANALYSIS", *BAREKENG: J. Math. & App.*, vol. 20, no. 2, pp. 1719-1730, Jun, 2026.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

The rapid advancement of digital technologies has led to the proliferation of massive amounts of textual data, particularly from social media platforms, customer reviews, and online news articles. Opinions are central to many aspects of human life as they directly influence everyday behavior and decision-making. When making a decision, individuals often consider various perspectives and opinions from others to gain a deeper understanding [1]. One analytical approach used in data processing is machine learning, which includes sentiment analysis as a specific application. Sentiment analysis is a component of Natural Language Processing (NLP) that uncovers opinions embedded in textual data [2].

Sentiment analysis is required to examine customer reviews, social media comments, and public opinion, which are now valuable sources of information for various purposes, including marketing, politics, and finance [3] One commonly used method for extracting insights from text is sentiment analysis, which is an automated process for determining whether a text has a positive, negative, or neutral sentiment [4].

Sentiment analysis is a computational technique aimed at identifying and classifying opinions expressed within textual data [1]. Early approaches relied on traditional machine learning models such as Support Vector Machine (SVM), Naïve Bayes, and Random Forest. Although these methods provided initial progress, they exhibited significant limitations in capturing contextual dependencies within text. The subsequent introduction of word embedding techniques, such as Word2Vec and GloVe, improved representation by mapping words to dense vector spaces; however, these methods still fell short of fully capturing contextual semantics.

One of the most widely used NLP models is Bidirectional Encoder Representations from Transformers (BERT), which was introduced by [2]. Although BERT has the advantage of understanding the meaning of a word in its sentence context, the model is less than optimal at handling long-term dependencies in text [5]. To address this limitation, the Long Short-Term Memory (LSTM) architecture is often applied to capture sequential patterns in textual data.

LSTM is an advanced variant of the Recurrent Neural Network (RNN) architecture, specifically designed to retain and manage information over long sequential intervals. It employs specialized gating mechanisms, namely the input gate, the forget gate, and the output gate, which collectively enhance its ability to model temporal dependencies and address challenges such as vanishing gradients in sequential data processing [6].

To enhance model performance, BERT is frequently integrated with advanced neural architectures such as BiLSTM. Empirical studies have demonstrated that this hybrid approach yields notable improvements in accuracy across multiple tasks [7][8].

A study conducted by [9] applied a combination of BERT and Bidirectional LSTM models to enhance the effectiveness and precision of sentiment classification processes. While BERT demonstrates strong capabilities in modeling contextual relationships among words, BiLSTM effectively captures long-term dependencies within sequential data. The combination of these two models enables a more comprehensive sentiment analysis approach by utilizing context understanding as well as time sequence learning [10]. BERT and BiLSTM are expected to complement each other in addressing the limitations of single models in handling long text sequences. Based on this, this study explores the hybrid integration of BERT and BiLSTM models for sentiment analysis, aiming to present an analytical overview and comparative insight into their combined architecture.

2. RESEARCH METHODS

This research framework is systematically developed based on three essential elements, consisting of the Visualization of Similarities Viewer (VosViewer), BERT, and BiLSTM.

2.1 Bibliometric with Visualization of Similarities Viewer (VosViewer)

The VosViewer is an open-access software designed to facilitate the analysis and visualization of bibliometric network maps derived from bibliographic data [11]. The strength of VosViewer compared to similar software tools is its ability to apply text mining for discovering relationships among phrases and to employ clustering techniques for testing and analyzing datasets [12]. VosViewer provides features that facilitate the presentation of bibliometric data through effective visualizations. It operates by integrating multiple documents to interpret their content, uncover recurring patterns, and identify emerging trends as written by [13], [14], [15]. In addition, VosViewer offers various interactive tools and customizable options that enhance user experience in navigating and exploring bibliometric data more efficiently [16][17].

Bibliometric data were collected from Scopus-indexed journal publications published between 2020 and 2025. A total of 500 documents were identified using the keywords “BERT”, “BiLSTM”, and “sentiment analysis”. The inclusion criteria were restricted to journal articles, conference papers, and book chapters to ensure the academic quality of the dataset. The bibliometric mapping of BERT and BiLSTM was generated using VOSviewer, and the resulting network visualization is shown in Figure 1.

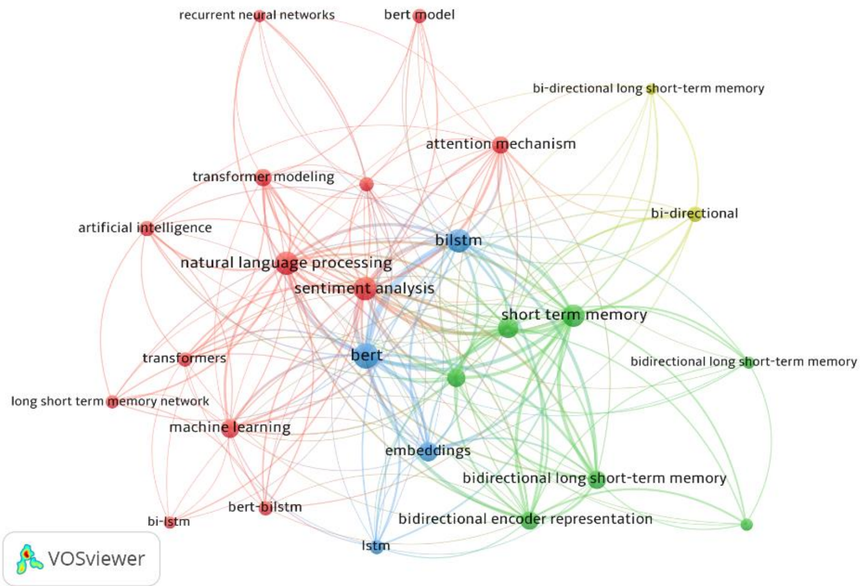

Figure 1. Visualization Restricted to Titles and Abstracts

Fig. 1 depicts the co-occurrence network of author keywords associated with BERT and BiLSTM in sentiment analysis research. Each node represents a keyword, with the node's size indicating its frequency of occurrence, and the connecting lines reflect co-occurrence relationships among terms. The color of each node corresponds to a specific cluster, signifying groups of related research themes. The visualization demonstrates that “sentiment analysis”, “BERT”, and “BiLSTM” occupy central positions within the network, emphasizing their dominant influence and strong interconnections in recent studies focusing on semantic representation and sequential text modeling.

Furthermore, the visualization identifies four distinct thematic clusters, each representing a different research direction. The red cluster includes terms such as “transformer modeling”, “attention mechanism”, and “natural language processing”, indicating a concentration of studies on transformer-based architectures and contextual representation learning. The green cluster contains terms like “short-term memory”, “bidirectional long short-term memory”, and “embeddings”, which emphasize recurrent neural network models used to capture sequential dependencies in textual data. The blue cluster links “BERT”, “BiLSTM”, and “sentiment analysis”, serving as an analytical bridge between transformer and recurrent architectures, thus highlighting the emerging focus on hybrid models that integrate contextual and sequential learning. Finally, the yellow cluster includes terms such as “bi-directional” and “bidirectional encoder representation”, reflecting methodological enhancements aimed at improving model efficiency and performance.

Despite the interconnected nature of these clusters, the visualization reveals that most prior studies have focused on transformer-based or recurrent architectures independently, with relatively limited attention to their systematic integration. This finding underscores a clear research gap while simultaneously validating the analytical direction of this study, which explores the integration of BERT and BiLSTM as a hybrid framework for sentiment analysis. Consequently, the bibliometric evidence not only supports the increasing convergence between contextual and sequential modeling but also indicates promising future research directions for optimizing hybrid architectures, advancing cross-lingual sentiment analysis, and developing computationally efficient integration between transformer and recurrent neural network models.

2.2 BERT Model

BERT is a model that uses the transformer's attention mechanism to understand the contextual relationships between words. The Transformer architecture consists of two main components: the encoder processes the input sequence, while the decoder generates predictions for specific tasks. The Transformer encoder integrates three essential components, namely residual connections, layer normalization, and feed-forward propagation, which are constructed upon the multi-head self-attention mechanism to enhance the representation learning process. Residual connections combined with layer normalization facilitate network training and effectively mitigate the vanishing gradient problem. Furthermore, the Transformer encoder represents each element of the input sequence as an embedding vector and incorporates positional information by adding a corresponding positional embedding of the same dimensionality, thereby enabling the model to capture the order of tokens within the sequence [18]. Unlike the directional model that reads the text sequentially, the encoder in the transformer is able to analyze all words simultaneously, so it is non-directional [2].

The base BERT model consists of 12 stacked encoder layers, each containing 12 self-attention heads and a feed-forward network with 768 hidden units. The final output, used as input for downstream tasks, generates high-quality word representations [2]. This model has been widely adopted to improve various NLP tasks, including sentiment classification, token segmentation, question answering, and named-entity identification. BERT is introduced to optimize pre-training by incorporating advanced semantic representations during the process.

The architectural design of BERT is shown in Fig. 2 [19]. For the trained model to be effectively utilized, the input data must be transformed into the proper format. Once sentence embeddings are obtained, the encoder layers process them as token sequences, incorporating the corresponding attention masks. The output maintains the same number of embeddings, each with a dimensionality equal to the hidden size.

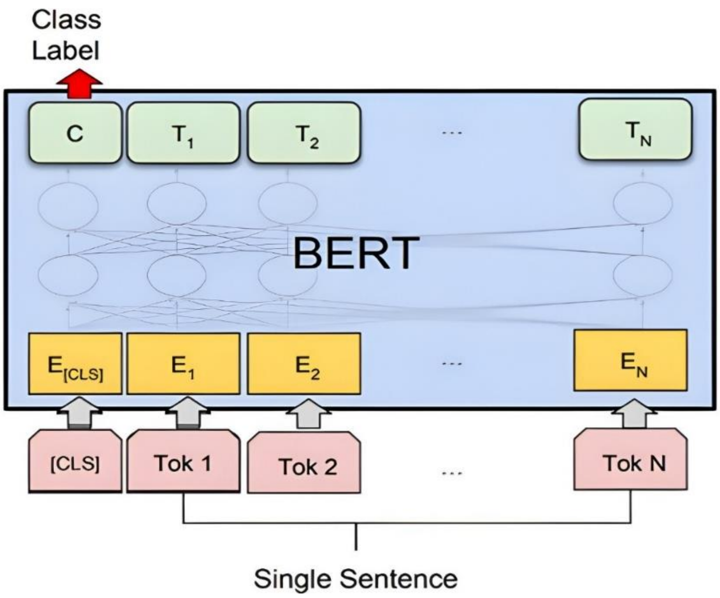

Figure 2. The Architectural Structure of BERT

For the classification task, a representative vector of the input sentence is used, derived from the model's [CLS] token hidden state. This vector is subsequently passed through a fully connected (linear) layer to generate sentiment predictions, consistent with the standard fine-tuning procedure for the BERT model in classification tasks [2].

2.3 BiLSTM Algorithm

2.3.1 LSTM Model

RNN is a type of artificial neural network composed of interconnected, interactive neurons, with connections represented by weighted arcs (W) that determine the strength of the relationship. This kind of network is particularly advantageous for handling inputs of different lengths and for processing sequential data, such as time series. It is widely used in areas such as machine translation, speech recognition, and pattern detection. In this type of artificial neural network, information propagates bidirectionally, allowing the model to preserve data sequences. Owing to its internal memory and recurrent loop structure, it can establish connections across long input sequences [20].

LSTM is a specialized variant of RNN introduced by [6] designed to retain and manage long-term dependencies within sequence-based data. The chain-like structure of LSTM is similar to that of RNNs, but the basic modules that make up LSTM have significant structural differences compared to those of other RNNs. RNNs are very effective in learning small sequences of data and have an advantage in this area [21].

The LSTM architecture is composed of recurrent modules, called cells, each comprising four interconnected neural networks. Within this structure, each cell transmits two forms of information to the next cell: the cell state and the hidden state. LSTM uses memory cells to maintain information throughout the learning process.

During computation, the input gate, forget gate, and output gate work collaboratively to regulate and manage the cell's memory. The input gate is essential for integrating pertinent information into the present cell state, utilizing a combination of sigmoid and tanh activation functions to regulate the input flow. The forget gate eliminates information deemed irrelevant by the LSTM, using a sigmoid activation function to selectively decide which data to preserve or discard. The output gate determines how much information from the current cell state to reveal, using a sigmoid activation function to control the flow of output data. The structure of the LSTM model, as referenced in [22], is presented in Fig. 3.

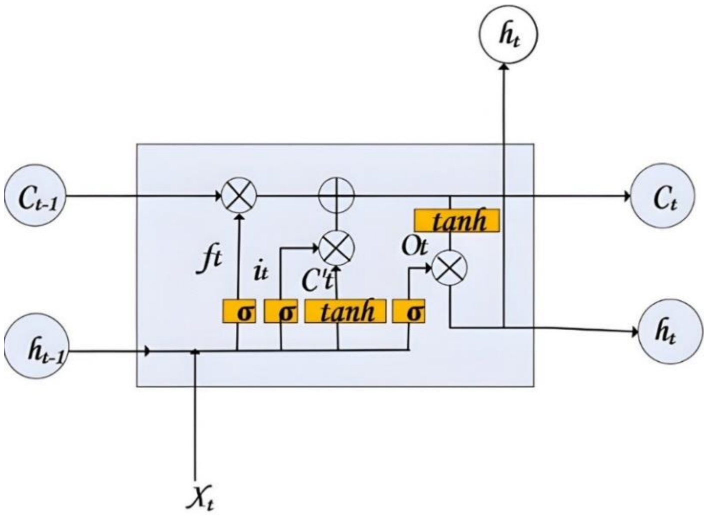

Figure 3. The Architectural Structure of LSTM

The initial stage of the LSTM computation involves determining which portions of the information stored in the previous cell state (c_{t-1}) should be discarded to optimize the learning process. This decision-making process is controlled by the forget gate (f_t) can be written as Eq. (1).

$$f_t = \sigma\big(W_{(f)}x_t + C_{(f)}h_{t-1} + b_{(f)}\big). \tag{1}$$

The subsequent step involves deciding which information should be added to the current cell state (c_t). This process includes two components: identifying the update values via the input gate (i_t) can be written as Eq. (2), and generating a candidate cell state vector (c'_t) can be written as Eq. (3).

$$i_t = \sigma\big(W_{(i)}x_t + C_{(i)}h_{t-1} + b_{(i)}\big), \tag{2}$$

$$c'_t = tanh\big(W_{(c)}x_t + C_{(c)}h_{t-1} + b_{(c)}\big). \tag{3}$$

Cell state (c_{t-1}) is updated to the new cell state (c_t) by intergrating the output of the forget gate, input gate, and the candidate vector (c'_t) can be written as Eq. (4).

$$c_t = f_t \cdot c_{t-1} + i_t \cdot c'_t. \tag{4}$$

The final stage involves generating the output using the updated cell state. This is performed through the output gate (o_t), as presented in Eq. (5). Subsequently, the output gate vector (o_t) according to [23] is combined with the (c_t) to calculate the hidden state (h_t), as shown in Eq. (6).

$$o_t = \sigma\left(W_{(o)}x_t + C_{(o)}h_{t-1} + b_{(o)}\right), \tag{5}$$

$$h_t = o_t * tanh(c_t), \tag{6}$$

where W and C is the weight matrix, x_t represents the input at time t and $b_{\{f,i,c,o\}}$ represents the set of bias vectors as written by [24].

2.3.2 BiLSTM Model

To enhance LSTM's ability to learn sequential patterns from both preceding and succeeding contexts, the BiLSTM architecture is used. The architecture comprises two parallel LSTM layers: one processes the input sequence in the forward temporal direction, capturing dependencies from past to future, while the other operates in the reverse direction, learning information from future to past. BiLSTM is a variant of RNN designed to capture long-term dependencies within sequential data [25]. This mechanism enables the model to capture contextual information from both preceding and succeeding elements within the encoded text.

In a standard LSTM, a single directional flow updates its hidden state based only on preceding data, without incorporating information from subsequent steps. To address this limitation, the BiLSTM architecture is configured to analyze input sequences bidirectionally. This configuration enables the model to extract meaningful contextual representations by considering both prior and subsequent elements in the sequence during training.

The BiLSTM architecture comprises two separate LSTM layers, with one processing the input sequence from left to right (forward) and the other from right to left (backward) [25]. The forward layer processes the input sequence from left to right by integrating the current input vector (x) with the previous hidden state (h_{t-1}). As shown in [19], the architectural design of the BiLSTM model is illustrated in Fig. 4.

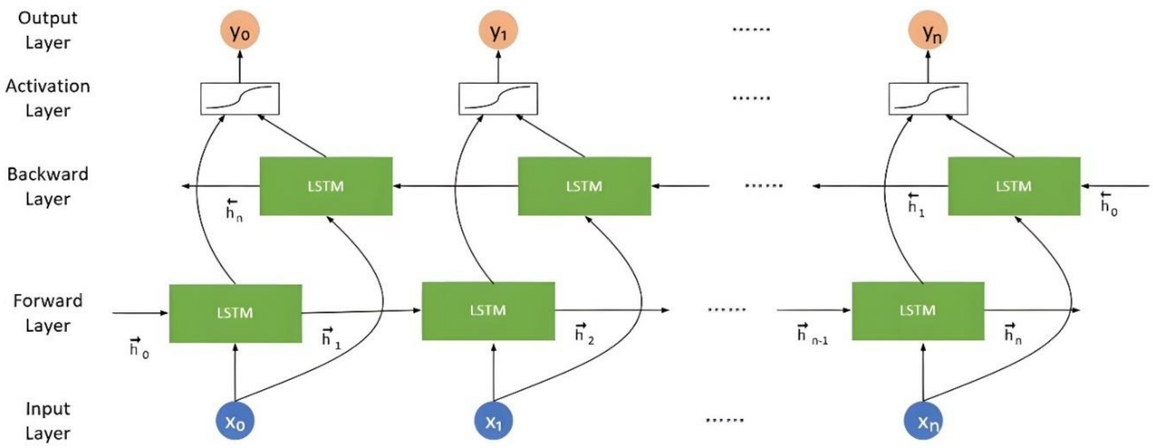

Figure 4. The Architectural Structure of BiLSTM

The output of the BiLSTM model is derived by integrating the hidden states produced by both the forward and backward LSTM layers. Let $\vec{h}_t$ and $\overleftarrow{h}_t$ denote the hidden state at time t corresponding to the forward and backward passes, respectively. The formulation of the BiLSTM output can be seen in Eqs. (7) and (8).

$$\vec{h}_t = LSTM\left(x_t, \vec{h}_{t-1}\right), \tag{7}$$

$$\overleftarrow{h}_t = LSTM(x_t, \overleftarrow{h}_{t-1}), \tag{8}$$

where:

x_t : the input vector at time t;

$\vec{h}_{t-1}$: the forward h_t from the previous time step $(t-1)$;

$\overleftarrow{h}_{t-1}$: the backward h_t from the subsequent time step $(t-1)$.

3. RESULTS AND DISCUSSION

This study presents an analytical discussion of the integration of BERT and BiLSTM models for sentiment analysis using advanced neural network frameworks. The BERT model transforms an input text sequence into contextual embeddings by leveraging a bidirectional Transformer architecture that integrates information from both preceding and succeeding elements of the sequence. BiLSTM is applied to capture long-term sequential dependencies from both directions, thereby complementing BERT's contextual representation.

Within such a framework, sentiment prediction is typically obtained using a Softmax activation function, which maps the output to a probability distribution over predefined sentiment categories. From an analytical standpoint, integrating BERT with BiLSTM is anticipated to provide a more comprehensive feature representation and to enhance the effectiveness of sentiment classification. The combined architecture of the BERT–BiLSTM model is illustrated in Fig. 5.

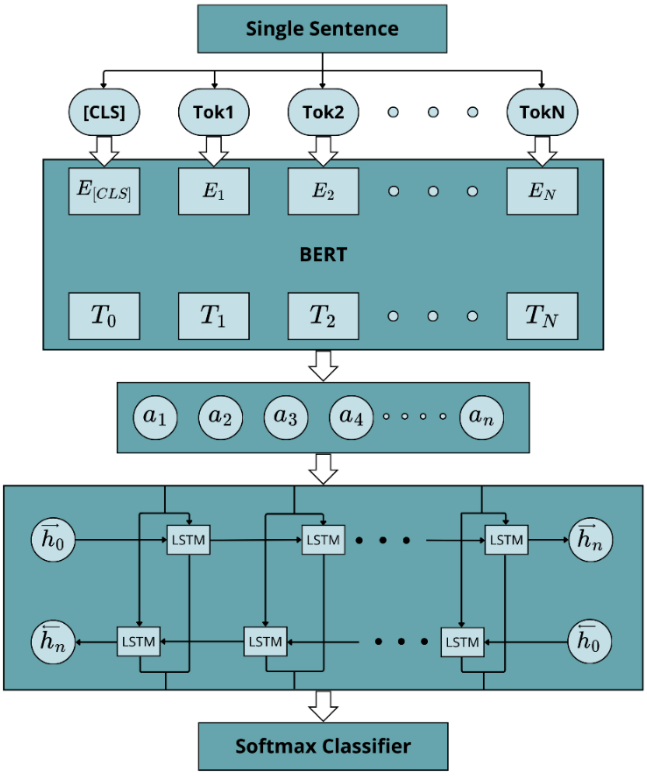

Figure 5. The Architectural Structure of BERT-BiLSTM

At the initial phase, the input text is processed by the BERT model, which converts each token into a contextual vector representation based on the surrounding context in the sequence. Because it works bidirectionally, BERT can capture the semantic meaning of a word by considering both its preceding and succeeding context within a sentence. In Fig. 5, $E_1, E_2, \dots, E_N$ represent the input word embeddings processed by the BERT layer, while $E_{[CLS]}$ denotes the special classification token used to indicate the beginning of the sentence, the vector T_0 corresponds to the output representation of the $[CLS]$ token after BERT training, and $T_1, T_2, \dots, T_N$ refer to the output vectors for each individual word generated by the BERT layer.

The BiLSTM layer functions bidirectionally, consisting of a forward pass that processes the input sequence from the beginning to the end, and a backward pass that processes it in reverse from the end to the beginning. This layer allows the model to understand long-term sequential patterns in text data. Subsequently, the Bi-LSTM model is employed to extract contextual features from the input sequence, where $\{h_0 \rightarrow h_1 \rightarrow h_2 \rightarrow \dots \rightarrow h_n\}$ represents the hidden layer sequence generated for the forward LSTM, $\{h_n \rightarrow \dots \rightarrow h_2 \rightarrow h_1 \rightarrow h_0\}$ the hidden layer sequence generated for the backward LSTM within the BiLSTM architecture. The results from these two directions are combined to form a more informative final feature representation.

In the integrated BERT–BiLSTM framework, the output vector $C = \{T_0, T_1, T_2, \dots, T_N\} \in R^n$ which include $[CLS]$ token from the BERT hidden layer is multiplied by a weight matrix $W_a \in R^{d_a \times n}$. The output produced from the previous layer is then input into the BiLSTM layer, which can be represented as Eq. (9).

$$a_t = g_1(W_a C_i + b_a), \tag{9}$$

with $1 \le t \le n$, n represents the dimensionality of the feature vector produced after fine-tuning the BERT model. The vector $a_t \in R^{d_a}$ serves as the input to the BiLSTM layer, while b_a denotes the bias vector with dimension d_a, and function g_1 is employed as the activation mechanism and is defined as the Sigmoid function. By combining Eqs. (7) and (8), the formulation of the output vector (v_t), can be represented as follows:

$$v_t = \vec{h}_t + \overleftarrow{h}_t, \tag{10}$$

where $\vec{h}_t \in R^{d_h}$, $\overleftarrow{h}_t \in R^{d_h}$.

In the standard BiLSTM formulation, the hidden state at time t can be represented by Eq. (11):

$$h_t^d = g_2\left(W_h^d a_t + C h_{t-1}^d + b_h^d\right), \tag{11}$$

where g_2 denotes the $Tanh$ function, and $W_h^d a_t \in R^{d_h \times d_a}$ corresponds to the weight matrix associated with the input vector (a_t), where $d \in \{0, 1\}$ indicates the two directions of the hidden states in the BiLSTM layer (forward and backward). The matrix C refers to the weight parameters applied to the output sequence of the hidden state (h^d) at time (t). The term h_{t-1}^d is the hidden state output from the previous time $(t-1)$ in direction d, while b_h represents the bias vector corresponding to the same directional component. Subsequently, the output sequences (h_d) from both directional hidden layers are concatenated to form the final feature vector H, which can be interpreted as representing contextual information at the sentence level. The fully connected layer then typically applies a Softmax function to map the feature vector H into class probabilities for sentiment prediction, as represented by Eq. (12).

$$p(y|H, W_h, b_h) = softmax(W_h H + b_h), \tag{12}$$

where:

$W_h \in R^{|h| \times |l|}$: the weight matrix of the output layer;

$b_h \in R^{|l|}$: the bias vector;

$|l|$: the number of target class.

To further contextualize the analytical framework outlined in this study, it is necessary to review how BERT, BiLSTM, and their integration have been applied in prior scholarly works. Accordingly, Table 1 presents a comparative synthesis of relevant studies, highlighting the principal contributions, reported strengths, and identified limitations of each model.

Table 1. Comparative Overview of BERT, BiLSTM, and BERT–BiLSTM

Model	Dataset	Metric	Result	Strengths	Limitations
BERT [2]	SST-2 (GLUE benchmark)	Accuracy	93.5%	Strong contextual embeddings	Limited handling of long sequences
BERT–BiLSTM [26]	IMDB Movie Reviews	Accuracy	97.67%	Better classification performance than BERT alone	Higher computational cost
LSTM, BiLSTM, BERT, RoBERTa, BERTweet [27]	Yelp Reviews	Accuracy/F1	LSTM 77%, Bi-LSTM 75%, BERT 67%, BERTweet 60%, RoBERTa 59%	Provides comparative benchmark across models	Transformer models underperformed on Yelp
BERT–BiLSTM–Attention [28]	Chinese Stock Reviews	Accuracy, F1	93.98%	Accuracy improved with attention mechanism	More complex architecture

Based on the comparative findings presented in Table 1, it can be concluded that prior studies consistently demonstrate the complementary strengths of BERT and BiLSTM. BERT achieves high accuracy on context-oriented benchmarks, such as SST-2 (93.5%), yet its performance tends to decline when applied to datasets with longer or more complex textual structures, as observed in Yelp Reviews, where recurrent models such as LSTM and BiLSTM outperform transformers. Conversely, BiLSTM exhibits strong capabilities in modeling sequential dependencies but lacks the deep contextual representation offered by BERT. Hybrid models, such as BERT–BiLSTM, which achieved 97.67% accuracy on the IMDB dataset, highlight the potential of integrating contextual embeddings with sequential learning, albeit at the cost of increased computational complexity and susceptibility to overfitting. These findings emphasize the importance of considering both dataset characteristics and computational trade-offs when selecting or designing sentiment analysis models. In this regard, the present study contributes by offering an analytical synthesis that integrates empirical evidence with methodological insights, positioning the BERT–BiLSTM integration as a promising framework for multilingual and domain-specific sentiment analysis. Furthermore, this framework underscores the systematic combination of contextual understanding and sequential modeling, thereby opening pathways for the development of more adaptive and efficient sentiment analysis approaches in future research.

4. CONCLUSION

This study presents an analytical synthesis of sentiment analysis research focusing on the integration of BERT and BiLSTM architectures. The bibliometric analysis using VosViewer revealed four dominant thematic clusters, indicating that most prior studies examined transformer-based and recurrent models independently, with limited exploration of their hybridization. The comparative findings from benchmark datasets demonstrate that combining BERT's contextual embeddings with BiLSTM's sequential learning consistently enhances sentiment classification accuracy, as evidenced by studies reporting over 97% accuracy on the IMDB dataset. However, this improvement is accompanied by higher computational costs and architectural complexity. The integrated framework discussed in this study contributes by reinforcing the connection between contextual and sequential representations as a foundation for developing more adaptive sentiment analysis systems, particularly in multilingual and domain-specific contexts. Future research should emphasize optimizing computational efficiency, exploring multilingual adaptability, and extending hybrid architectures to broader text analytics applications.

Author Contributions

Nicolas Ray Amarco Tambunan: Conceptualization, Methodology, Writing-Original Draft Preparation, Visualization. Dewi Retno Sari Saputro: Conceptualization, Writing-Review and Editing, Visualization, Supervision. Purnami Widyaningsih: Writing-Review and Editing, Resource, Supervision. All authors discussed the results and contributed to the final manuscript.

Funding Statement

This research was financially supported by the Institute for Research and Community Service (LPPM), Universitas Sebelas Maret (UNS), through the research group grant based on the approval letter for non-APBN-funded research No. 370/UN27.22/PT.01.03/2025.

Acknowledgment

The authors sincerely acknowledge the Softcomputing Mathematics Research Group for their continuous encouragement and valuable suggestions, which greatly contributed to the improvement of this paper.

Declarations

The authors declare no conflicts of interest to report for this.

Declaration of Generative AI and AI-assisted technologies

Generative AI tools (e.g., ChatGPT) were used solely for language refinement (grammar, spelling, and clarity). The scientific content, analysis, interpretation, and conclusions were developed entirely by the authors. The authors reviewed and approved all final text.

REFERENCES

[1] B. Liu, "SENTIMENT ANALYSIS AND OPINION MINING," *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1–167, May 2012, doi: https://doi.org/10.2200/S00416ED1V01Y201204HLT016.

[2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: PRE-TRAINING OF DEEP BIDIRECTIONAL TRANSFORMERS FOR LANGUAGE UNDERSTANDING," May 2019, doi: https://doi.org/10.18653/v1/N19-1423.

[3] B. Liu, "SENTIMENT ANALYSIS: MINING OPINIONS, SENTIMENTS, AND EMOTIONS, SECOND EDITION," *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Second Edition*, no. May, pp. 1–432, 2020, doi: https://doi.org/10.1017/9781108639286.

[4] W. Medhat, A. Hassan, and H. Korashy, "SENTIMENT ANALYSIS ALGORITHMS AND APPLICATIONS: A SURVEY," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, 2014, doi: https://doi.org/10.1016/j.asej.2014.04.011

[5] Z. Huang, W. Xu, and K. Yu, "BIDIRECTIONAL LSTM-CRF MODELS FOR SEQUENCE TAGGING," 2015, doi: https://doi.org/10.48550/arXiv.1508.01991.

[6] S. Hochreiter and J. Schmidhuber, "LONG SHORT-TERM MEMORY," *Neural Comput*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: https://doi.org/10.1162/neco.1997.9.8.1735.

[7] A. B. Alawi and F. Bozkurt, "A HYBRID MACHINE LEARNING MODEL FOR SENTIMENT ANALYSIS AND SATISFACTION ASSESSMENT WITH TURKISH UNIVERSITIES USING TWITTER DATA," *Decision Analytics Journal*, vol. 11, no. April, p. 100473, 2024, doi: https://doi.org/10.1016/j.dajour.2024.100473.

[8] R. E. N. Cai, B. I. N. Qin, Y. Chen, S. Chen, and W. E. I. Wang, "SENTIMENT ANALYSIS ABOUT INVESTORS AND CONSUMERS IN ENERGY MARKET BASED ON BERT-BILSTM," vol. 8, 2020, doi: https://doi.org/10.1109/ACCESS.2020.3024750.

[9] X. Zhou, "SENTIMENT ANALYSIS OF THE CONSUMER REVIEW TEXT BASED ON BERT-BILSTM IN A SOCIAL MEDIA ENVIRONMENT," *International Journal of Information Technologies and Systems Approach*, vol. 16, pp. 1–16, Jan. 2023, doi: https://doi.org/10.4018/IJITSA.325618.

[10] M. V. Joseph, "A BI-LSTM AND GRU HYBRID NEURAL NETWORK WITH BERT FEATURE EXTRACTION FOR AMAZON TEXTUAL REVIEW ANALYSIS," *International Journal of Engineering Trends and Technology*, vol. 70, no. 5, pp. 131–144, 2022, doi: https://doi.org/10.14445/22315381/IJETT-V70I5P216.

[11] E. W. Rice, "BIBLIOMETRIC EVALUATIONS OF MODERN CLINICAL CHEMISTRY ARE NEEDED.," *Clin Chem*, vol. 29, no. 10, pp. 1858–1859, Oct. 1983, doi: https://doi.org/10.1093/clinchem/29.10.1858.

[12] D. R. S. Saputro, H. Prasetyo, A. Wibowo, F. Khairina, K. Sidiq, and G. N. A. Wibowo, "BIBLIOMETRIC ANALYSIS OF NEURAL BASIS EXPANSION ANALYSIS FOR INTERPRETABLE TIME SERIES (N-BEATS) FOR RESEARCH TREND MAPPING," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 17, no. 2, pp. 1103–1112, Jun. 2023, doi: https://doi.org/10.30598/barekengvol17iss2pp1103-1112.

[13] D. F. Al Husaeni and A. B. D. Nandiyanto, "BIBLIOMETRIC USING VOSVIEWER WITH PUBLISH OR PERISH (USING GOOGLE SCHOLAR DATA): FROM STEP-BY-STEP PROCESSING FOR USERS TO THE PRACTICAL EXAMPLES IN THE ANALYSIS OF DIGITAL LEARNING ARTICLES IN PRE AND POST COVID-19 PANDEMIC," *ASEAN Journal of Science and Engineering*, vol. 2, no. 1, pp. 19–46, 2022, doi: https://doi.org/10.17509/ajse.v2i1.37368.

[14] A. Kirby, "EXPLORATORY BIBLIOMETRICS: USING VOSVIEWER AS A PRELIMINARY RESEARCH TOOL," 2023. doi: https://doi.org/10.3390/publications11010010.

[15] M. Dubyna, O. Popelo, N. Kholiavko, A. Zhavoronok, M. Fedyshyn, and I. Yakushko, "MAPPING THE LITERATURE ON FINANCIAL BEHAVIOR: A BIBLIOMETRIC ANALYSIS USING THE VOSVIEWER PROGRAM," *WSEAS Transactions on Business and Economics*, vol. 19, no. January, pp. 231–246, 2022, doi: https://doi.org/10.37394/23207.2022.19.22.

[16] L. Waltman, N. J. van Eck, and E. Noyons, "A UNIFIED APPROACH TO MAPPING AND CLUSTERING OF BIBLIOMETRIC NETWORKS," *J Informetr*, vol. 4, pp. 629–635, Jun. 2010, doi: https://doi.org/10.1016/j.joi.2010.07.002.

[17] N. J. van Eck and L. Waltman, "SOFTWARE SURVEY: VOSVIEWER, A COMPUTER PROGRAM FOR BIBLIOMETRIC MAPPING," *Scientometrics*, vol. 84, no. 2, pp. 523–538, 2010, doi: https://doi.org/10.1007/s11192-009-0146-3.

[18] H. Xu, Y. Huang, Y. Zhu, K. Audhkhasi, and B. Ramabhadran, "CONVOLUTIONAL DROPOUT AND WORDPIECE AUGMENTATION FOR END-TO-END SPEECH RECOGNITION", 2021, doi: https://doi.org/10.1109/ICASSP39728.2021.9415004.

[19] N. Rai, D. Kumar, N. Kaushik, C. Raj, and A. Ali, "FAKE NEWS CLASSIFICATION USING TRANSFORMER BASED ENHANCED LSTM AND BERT," *International Journal of Cognitive Computing in Engineering*, vol. 3, no. October 2021, pp. 98–105, 2022, doi: https://doi.org/10.1016/j.ijcce.2022.03.003.

[20] Y. Zheng and D. (Xuejun) Wang, "A SURVEY OF RECOMMENDER SYSTEMS WITH MULTI-OBJECTIVE OPTIMIZATION," *Neurocomputing*, vol. 474, pp. 141–153, Feb. 2022, doi: https://doi.org/10.1016/j.neucom.2021.11.041.

[21] M. Luengo and D. García-Marín, "THE PERFORMANCE OF TRUTH: POLITICIANS, FACT-CHECKING JOURNALISM, AND THE STRUGGLE TO TACKLE COVID-19 MISINFORMATION," *Am J Cult Sociol*, vol. 8, no. 3, pp. 405–427, 2020, doi: https://doi.org/10.1057/s41290-020-00115-w.

[22] D. Chi, T. Huang, Z. Jia, and S. Zhang, "RESEARCH ON SENTIMENT ANALYSIS OF HOTEL REVIEW TEXT BASED ON BERT-TCN-BILSTM-ATTENTION MODEL," *Array*, vol. 25, no. July 2024, p. 100378, 2025, doi: https://doi.org/10.1016/j.array.2025.100378.

[23] M. N. Abda, P. Khano, D. Retno, S. Saputro, and A. Wibowo, "SENTIMENT ANALYSIS WITH LONG-SHORT TERM MEMORY (LSTM) AND GATED RECURRENT UNIT (GRU) ALGORITHMS," vol. 17, no. 4, pp. 2235–2242, 2023, doi: https://doi.org/10.30598/barekengvol17iss4pp2235-2242.

[24] T. Ganegedara, *NATURAL LANGUAGE PROCESSING WITH TENSORFLOW: TEACH LANGUAGE TO MACHINES USING PYTHON'S DEEP LEARNING LIBRARY*. in Expert insight. Packt Publishing, 2018.

[25] P. Zhou, Z. Qi, S. Zheng, J. Xu, H. Bao, and B. Xu, "TEXT CLASSIFICATION IMPROVED BY INTEGRATING BIDIRECTIONAL LSTM WITH TWO-DIMENSIONAL MAX POOLING," *COLING 2016 - 26th International*

Conference on Computational Linguistics, Proceedings of COLING 2016: Technical Papers, vol. 2, no. 1, pp. 3485–3495, 2016, doi: https://doi.org/10.48550/arXiv.1611.06639.

[26] G. Nkhata, S. Gauch, U. Anjum, and J. Zhan, "FINE-TUNING BERT WITH BIDIRECTIONAL LSTM FOR FINE-GRAINED MOVIE REVIEWS SENTIMENT ANALYSIS," Feb. 2025, doi: https://doi.org/10.48550/arXiv.2502.20682.

[27] R. Belaroussi, S. C. Noufe, F. Dupin, and P. O. Vandanjon, "POLARITY OF YELP REVIEWS: A BERT–LSTM COMPARATIVE STUDY," *Big Data and Cognitive Computing*, vol. 9, no. 5, May 2025, doi: https://doi.org/10.3390/bdcc9050140.

[28] X. Li, L. Chen, B. Chen, and X. Ge, "BERT-BILSTM-ATTENTION MODEL FOR SENTIMENT ANALYSIS ON CHINESE STOCK REVIEWS," *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, Jan. 2024, doi: https://doi.org/10.2478/amns-2024-1847.