

Article

Not peer-reviewed version

---

# Enhancing Thai Sentiment Analysis with a Lightweight WangchanBERTa-CNN Hybrid and Optimal Kernel Configuration

---

[Panida Songram](#) , [Suchart Khummanee](#) , [Jirapond Muangprathub](#) , [Khanabhorn Kawattikul](#) \*

Posted Date: 9 April 2026

doi: 10.20944/preprints202604.0631.v1

Keywords: hybrid model; kernel size configuration; Thai sentiment analysis; convolutional neural networks; WangchanBERTa-CNN; Thai social media texts







Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Enhancing Thai Sentiment Analysis with a Lightweight WangchanBERTa–CNN Hybrid and Optimal Kernel Configuration

Panida Songram <sup>1</sup>, Suchart Khummanee <sup>1</sup>, Jirapond Muangprathub <sup>2</sup>  
and Khanabhorn Kawattikul <sup>3,\*</sup>

<sup>1</sup> Department of Computer Science, Faculty of Informatics, Maharakham University, Maharakham 44150, Thailand

<sup>2</sup> Department of Applied Mathematics and Informatics, Faculty of Science and Industrial Technology, Prince of Songkla University, Surat Thani Campus, Surat Thani 84000, Thailand

<sup>3</sup> Department of Information Technology, Faculty of Social Technology, Rajamangala University of Technology Tawan-Ok, Chanthaburi 22210, Thailand

\* Correspondence: khanabhorn\_ka@rmutto.ac.th

## Abstract

Sentiment analysis of Thai social media texts remains challenging due to the lack of explicit word boundaries, informal language, and high linguistic variability, further exacerbated by class imbalance, short texts, and ambiguous categories. Recent transformer-based models, particularly WangchanBERTa, have shown strong performance in Thai NLP; however, their ability to extract fine-grained, sentiment-specific local features remains limited when applied to noisy, short social media texts. This study proposes a lightweight hybrid framework that integrates WangchanBERTa with multiple neural architectures, including convolutional neural networks (CNN), bidirectional long short-term memory (BiLSTM), and bidirectional gated recurrent units (BiGRU). A comparative evaluation is conducted to assess their effectiveness for four-class Thai sentiment classification. In addition, the impact of CNN kernel size configurations on sentiment feature extraction is systematically investigated, an aspect that has received limited attention in prior Thai NLP research. Experiments conducted on the WISESIGHT benchmark demonstrate that the WangchanBERTa–CNN model with medium kernel sizes [2, 3, 4] achieves the best performance. It achieves a macro-average F1-score of 65.80%, outperforming the current state-of-the-art on the WISESIGHT benchmark by 4.94%. In addition, it achieves a macro-average F1-score of 94.11% on the class-balanced dataset. These results are achieved while maintaining a relatively low parameter count. These findings confirm that combining contextualized global embeddings with local n-gram feature extraction provides an effective and efficient solution for sentiment classification of short, noisy, and linguistically diverse Thai social media texts.

**Keywords:** hybrid model; kernel size configuration; Thai sentiment analysis; convolutional neural networks; WangchanBERTa–CNN; Thai social media texts

## 1. Introduction

Sentiment analysis is an important methodology for investigating public sentiment, tracking brand reputation, and deriving actionable insights from large volumes of social media reviews. Although effective techniques have been proposed for sentiment analysis in highly researched languages such as English and Chinese, sentiment analysis in low-resource languages remains a significant challenge, especially Thai. Thai is a tonal language that lacks explicit word boundaries and exhibits complex linguistic characteristics, making text processing inherently difficult. Therefore, specialized approaches are required for sentiment analysis in this language. Sentiment analysis on Thai social media texts, characterized by creative spelling variations, emoji usage, code-mixing with English, and unique internet slang, further complicates sentiment analysis on Thai texts. Many approaches have been

proposed for Thai sentiment analysis. Traditional lexicon-based sentiment analysis was first proposed, but it is ineffective at capturing the complexity of sentiment in social media text. Early machine learning-based sentiment analysis relied heavily on manually engineered features, which often lack generalizability. Consequently, deep learning approaches have emerged as a more effective solution. Various deep learning models have been proposed for sentiment analysis, such as LSTM, BiLSTM, CNN, GRU, and BiGRU. Among them, hybrid deep learning models were also proposed to enhance the performance of Thai sentiment analysis [1]. Hybrid models demonstrate significantly improved performance over single-architecture models.

Recently, transformer models have revolutionized natural language processing across many languages. Well-known representatives of this family include BERT [2], mBERT [3], RoBERTa [4], ALBERT [5], XLM-RoBERTa [6], and ELECTRA [7]. Some of these models include Thai language, such as mBERT, ALBERT, and XLM-RoBERTa. Unfortunately, previous studies have shown that models designed to handle multiple languages generally deliver weaker results than models trained and optimized for a single language [8–11]. For the Thai language, the development of WangchanBERTa [12], a RoBERTa-based model pre-trained on large-scale Thai corpora, has yielded powerful contextualized representations that capture semantic and syntactic information specific to Thai. However, while these pre-trained models excel at encoding linguistic knowledge, they may not optimally extract sentiment-specific features that are crucial for fine-grained sentiment classification tasks. The combination of the WangchanBERTa and deep learning models has thus been proposed and shown promising performance for sentiment analysis in Thai [13].

Previous studies have explored various deep learning architectures to improve sentiment analysis in Thai. However, sentiment analysis of Thai social media texts remains challenging, particularly on the WISESIGHT Sentiment benchmark dataset, due to class imbalance, short text length, noisy user-generated content, and ambiguity between sentiment categories such as neutral and question. In addition, existing studies have not systematically explored CNN kernel size configurations for sentiment-specific feature extraction in Thai. This paper proposes lightweight hybrid models that integrate WangchanBERTa with deep learning architectures, including BiLSTM, BiGRU, and CNN, to enhance performance while maintaining computational efficiency. In addition, kernel size configurations for CNN-based sentiment feature extraction are systematically evaluated to assess their impact on performance across different sentiment categories. Experimental results demonstrate that the WangchanBERTa-based hybrid model with optimal selection of CNN kernel size improves performance and outperforms the state-of-the-art in terms of macro-average F1-score. Based on these findings, the main contributions of this study are as follows:

1. A lightweight hybrid framework integrating WangchanBERTa with CNN for Thai sentiment analysis, balancing performance and computational efficiency.
2. A systematic analysis of CNN kernel size configurations for sentiment feature extraction, providing new insights into their impact across sentiment categories.
3. Empirical improvement over the state-of-the-art, achieving a 4.94% increase in macro-average F1-score on the WISESIGHT benchmark.
4. Demonstration of an effective combination of contextualized embeddings (global features) and local feature extraction for handling short, noisy, and imbalanced Thai social media texts.

## 2. Related Work

The field of sentiment analysis in Thai has developed substantially over the last decade and a half, from lexicon-based methods to deep learning architectures. Past studies sought to improve methods for understanding sentiment in the Thai context, while more recent studies have used single transformer-based models and hybrid transformer-based models combined with other deep learning models to deeply understand the Thai language. Previous sentiment analysis studies can be grouped into two categories as follows.

The first group emphasized lexicon-based and machine learning methods. For example, Phien-thrakul et al. explored SVM with multiple kernel functions for sentiment classification in Thai text [14]. The combination of different kernel functions and a single kernel function was evaluated, and it was reported that the combination could capture diverse linguistic features and improve classification performance compared to single-kernel approaches on product reviews. Lertsuksakda et al. proposed a novel approach to constructing Thai sentiment terms using the hourglass of emotions model [15]. They moved simple positive-negative polarity to capture more nuanced emotional states. Sentiment lexicons reflecting complex emotional dimensions were constructed. Chirawichitchai investigated emotion classification in Thai text using different term weighting and machine learning techniques [16]. The various term weighting schemes were explored to improve classification accuracy. The results showed that appropriately weighted features could significantly enhance the performance of machine learning models for Thai emotion classification. Chirawichitchai proposed a term weighting scheme based on term occurrence ratio optimized for sentiment analysis [17]. The term weighting scheme was proposed in mathematical formulations for calculating term importance that better reflected the sentiment characteristics of Thai words. The proposed weighting scheme improved classification performance by emphasizing terms with strong discriminative power for sentiment polarity.

Pasupa et al. applied SVMs to sentiment analysis of Thai children's stories [18]. They extended sentiment analysis typically performed in domains such as product reviews and social media. Netisopakul and Lertsuksakda employed hypothesis testing based on observations from Thai sentiment classification experiments [19]. A more rigorous statistical approach was used to evaluate sentiment classification methods, employing hypothesis testing to validate performance differences between approaches. Haruechaiyasak et al. introduced S-Sense, which is a comprehensive framework for sentiment analysis. The framework was specifically designed for social media sensing [20]. It integrated multiple components, including preprocessing, feature extraction, and classification, tailored to the informal and often abbreviated language used in Thai social media. Porntrakoon and Moemeng introduced the SenseComp (Sentiment Compensation) technique for multi-dimensional analysis of consumer reviews in Thai [21]. They observed that consumer reviews often contain mixed sentiments across different product dimensions. As a result, a more sophisticated analysis was required than simple overall polarity classification. To address this, the proposed SenseComp technique improved sentiment analysis by evaluating multiple attributes separately. Taemung and Chirawichitchai applied SVM to analyze sentiment in Thai product reviews [22]. They focused on e-commerce applications and addressed practical challenges in classifying Thai-language customer reviews. The study showed that SVM could achieve reasonable performance on commercial sentiment analysis tasks. However, it still struggled to handle complex or ambiguous sentiment expressions.

The second group focused on deep learning and transformer approaches. For example, Vateekul and Koomsubha conducted comprehensive studies by applying deep learning techniques on Thai Twitter data [23]. Various deep learning architectures, including DCNN (Dynamic Convolutional Neural Network) and LSTM, were explored for sentiment analysis of Thai social texts. The study showed that the deep learning approaches outperformed traditional machine learning methods on Thai social media sentiment analysis. The best model was DCNN, achieving the highest accuracy with 75.35%. Pasupa and Seneewong also proposed a comparative study of deep learning techniques for Thai sentiment analysis [24]. Various deep learning architectures were systematically evaluated, and the impact of different input representations, including word embeddings, POS tags, and sentiment features, was explored. The study showed that deep learning approaches work best for Thai sentiment analysis. The CNN architecture, combined with the three feature types, could achieve the highest F1 Score of 81.70%. Thong-iad and Netisopakul compared different methods for Thai sentence sentiment tagging using Thai sentiment resources [25]. Different sentiment lexicons and tagging approaches were evaluated for the classification task. The results showed that using adverb and adjective synsets alone could give the highest accuracy of emotion classification. Traditional machine learning, SVM, Random Forest, performs well, but deep learning models, CNN, LSTM, BERT-based, often surpass

them, especially with advanced techniques like hyperparameter tuning and incorporating linguistic features (POS, sentiment values). SVM offers strong performance among traditional methods, while hybrid deep learning models and transformer models (like WangchanBERTa) achieve higher accuracy

Lowphansirikul et al. developed WangchanBERTa, a transformer-based pre-trained language model specifically for Thai [12]. WangchanBERTa consistently outperforms established multilingual models like mBERT and XLM-R across a variety of benchmarks, achieving in NER (Named Entity Recognition), sentiment analysis, and POS tagging. WangchanBERTa has since become a foundation model for numerous Thai NLP tasks, including sentiment analysis. Pasupa and Seneewong proposed hybrid deep learning models that combined multiple neural architectures for Thai sentiment analysis [1]. The hybrid model, integrating CNNs and LSTMs, could capture both local features and long-range dependencies. It achieved higher performance than single-architecture models. From the proposed hybrid models, Bi-LSTM-CNN achieved the highest performance. It achieved macro-average F1-scores of 74.36% on ThaiTales, 77.07% on ThaiEconTwitter, and 55.21% on the WISESIGHT datasets. Jitboonyapinit et al. investigated sentiment analysis on Thai social media using convolutional neural networks combined with long short-term memory networks [26]. The challenges of this work are about informal language and unique linguistic patterns in Thai social media posts. The CNN-LSTM model successfully extracted spatial and temporal features. It achieved 85.00% accuracy of product reviews on social media. Khamphakdee and Seresangtakul developed an efficient deep learning approach optimized for Thai sentiment analysis [27]. This work focused on balancing model performance with computational efficiency. Modifications to the model architecture reduced training time while maintaining high accuracy.

Nokkaew et al. analyzed online public opinion regarding major infrastructure projects using advanced machine learning and deep learning for Thai sentiment analysis [28]. Their research demonstrated the application of sentiment analysis to policy-relevant social issues by analyzing public discourse on the Thailand-China high-speed train and Laos-China railway projects. Comment sentiment classification was performed using six approaches: linear regression, naive bayes, random forest, Bi-LSTM, BERT-Base-Thai, and WangchanBERTa. The WangchanBERTa model achieved 94.57% accuracy. Suraratchai and Phoomvuthisarn proposed a hybrid method combining WangchanBERTa with CNN and BiLSTM architectures for Thai sentiment analysis [13]. Their research built upon the pre-trained WangchanBERTa model by adding convolutional and bidirectional recurrent layers to capture additional linguistic features. The hybrid architecture achieved competitive performance on the WISESIGHT and the Thai Children's Tales datasets. The Parallel Hybrid approach, WangchanBERTa-CNN-BiLSTM, achieved the highest macro-average F1-scores, reaching 62.70% on the WISESIGHT dataset and 78.59% on the Thai Children's Tales dataset. Satjathanakul and Siriborvornratanakul focused on improving sentiment polarity classification on the Thai product reviews dataset using modern Transformer-based architectures [29]. The fine-tuned WangchanBERTa model achieved performance metrics ranging between 66% and 93% on the product reviews dataset. Emphan et al. enhanced the performance of the sentiment analysis model using GridSearchCV for hyperparameter optimization. This work focused on classifying sentiment on electric vehicle discussions in Thailand [30]. Hyperparameter tuning was used to identify optimal configurations for Thai sentiment classification models. The study demonstrated that careful hyperparameter tuning could improve performance.

Previous work has demonstrated that transformer-based models, especially WangchanBERTa, achieve high performance in contextual understanding of Thai linguistic nuances. Furthermore, combining WangchanBERTa with other deep learning techniques could improve the performance of Thai sentiment analysis even on imbalanced datasets such as WISESIGHT. As in previous work, they extensively explored various hybrid models, including WangchanBERTa, with other deep learning models for Thai sentiment analysis, but the impact of kernel size selection when combining WangchanBERTa and CNN across different sentiment categories remains underexplored. Therefore, this paper proposes a hybrid architecture combining WangchanBERTa and CNN and systematically investigates optimal kernel configurations to enhance sentiment analysis performance on the im-

balanced WISESIGHT dataset. In addition, two alternative hybrid models, WangchanBERTa with BiGRU and WangchanBERTa with BiLSTM, are developed for comparative evaluation against the WangchanBERTa-CNN model.

### 3. Materials and Methods

Figure 1 shows an overview of the proposed hybrid deep learning framework for Thai sentiment analysis. In the pre-processing stage, two approaches were employed. In the first approach, we did not apply any imbalance-handling techniques to ensure a fair comparison with prior studies that used the original dataset without modification. In the second approach, a data balancing technique, namely random oversampling, was applied to address class imbalance and improve the representation of minority classes. The input texts are tokenized using the WangchanBERTa tokenizer, and special tokens ([CLS], [SEP]) are added. The tokens are then converted into token IDs. These IDs are fed to the hybrid deep learning models that integrate WangchanBERTa with CNN, BiLSTM, and BiGRU. The hybrid deep learning models are then evaluated to find the best-performing model. The details of each phase are described below.

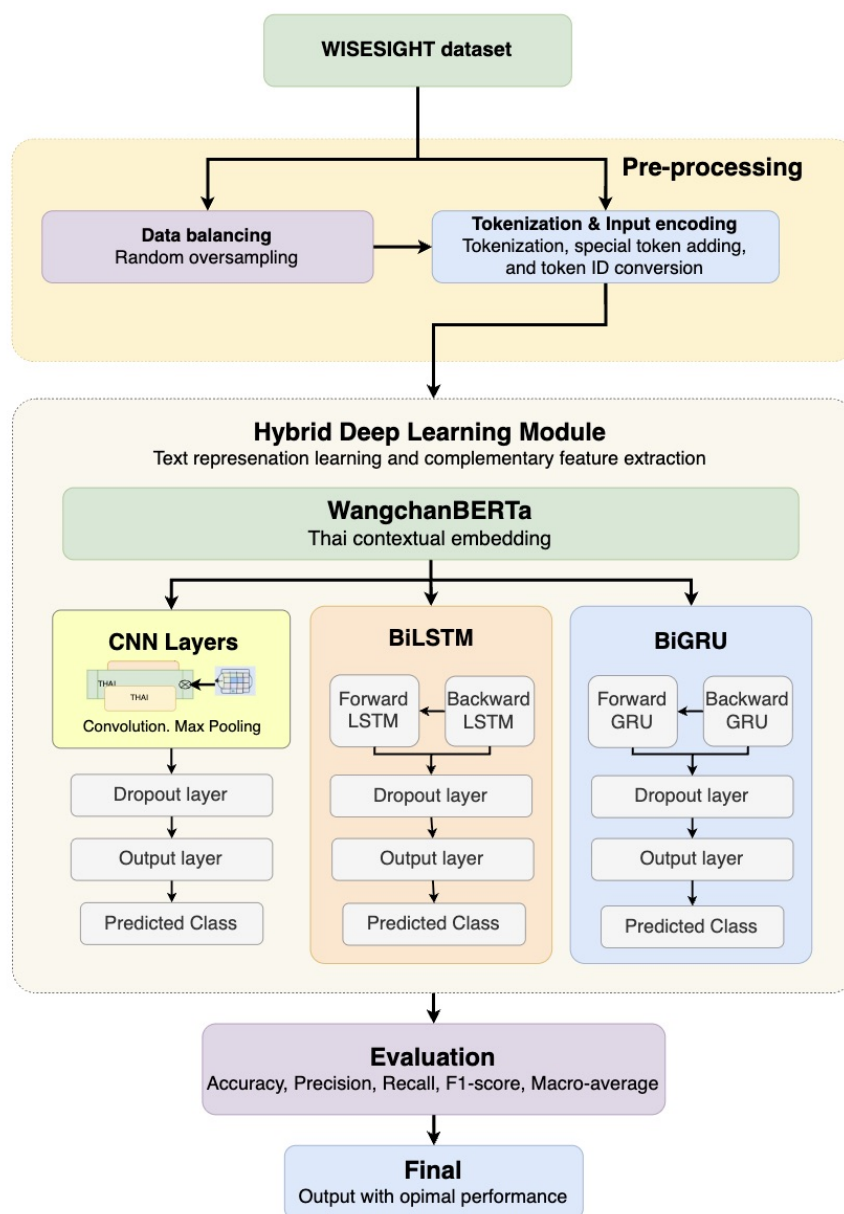


Figure 1. Overview of the proposed hybrid deep learning framework

### 3.1. The WISESIGHT Dataset

The WISESIGHT dataset has been widely used for Thai sentiment analysis in many studies. It was introduced by Suriyawongkul et al. [31]. The data was collected by WISESIGHT Thailand, one of the largest social media analysis companies that monitors social media discussions across many channels. The data comprises real messages written in Thai, collected from many popular social networking sites, including Facebook, Twitter, Instagram, YouTube, and Pantip. This publicly available dataset is available from UCI or Hugging Face. It contains social media posts annotated into four sentiment categories: positive, neutral, negative, and question. The WISESIGHT dataset is challenging for sentiment analysis competitions because it contains informal, noisy Thai social media text with slang, misspellings, and no clear word boundaries, making accurate tokenization and interpretation difficult. Additionally, short context on average (as shown in Figure 2), mixed sentiments, class imbalance, and subtle differences between categories (e.g., neutral vs. question) further complicate sentiment classification as shown in Figure 3. The WISESIGHT benchmark is split into training, validation, and testing sets for competition in prior studies, as shown in Figure 4 and Table 1.

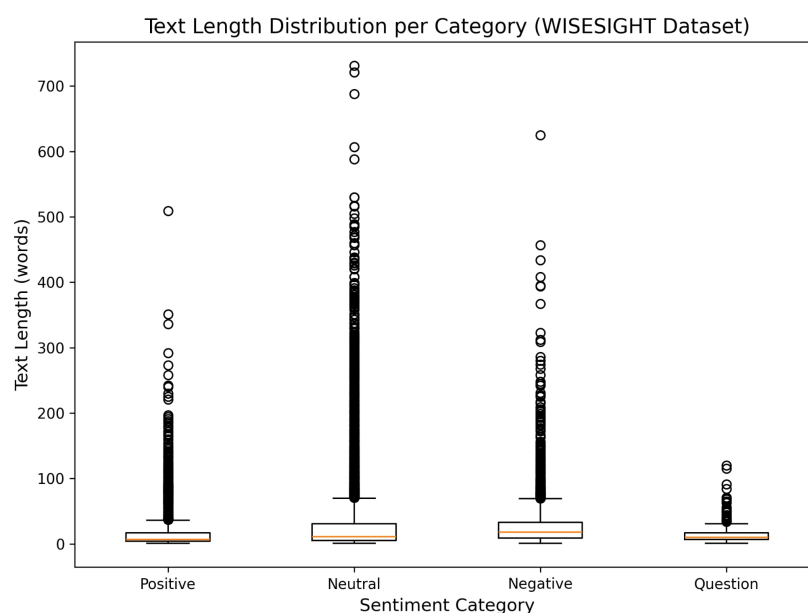


Figure 2. Text length distribution per category

Class Distribution of Wisightsentiment Dataset

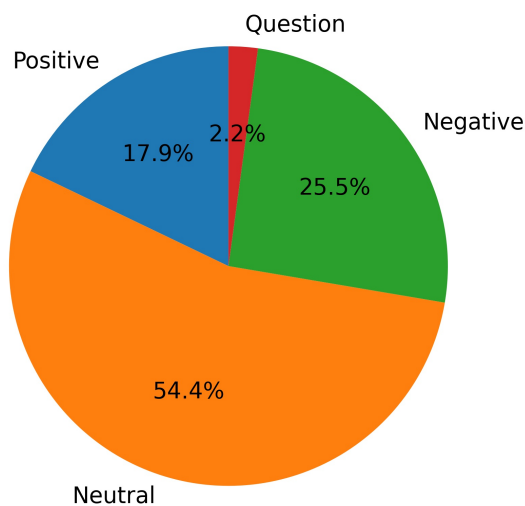


Figure 3. Class distribution the WISESIGHT dataset

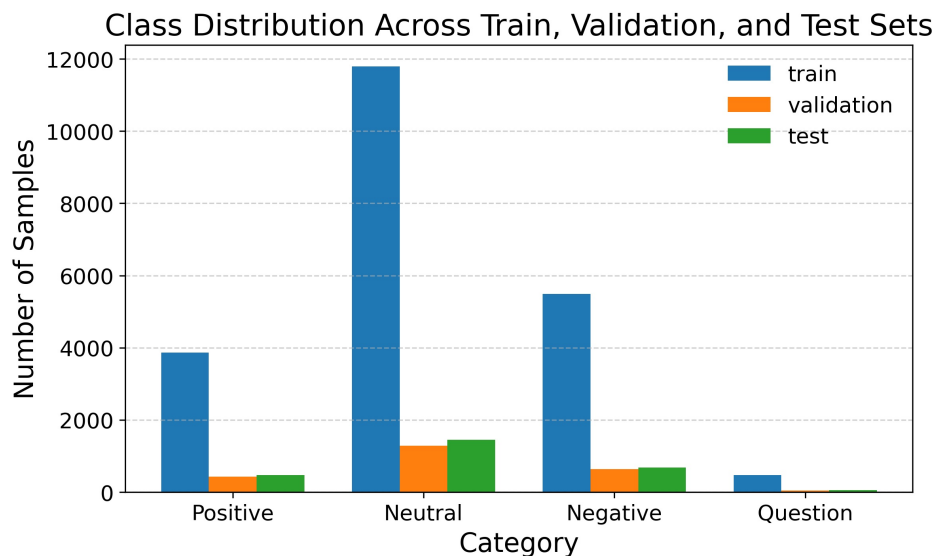


Figure 4. Class distribution across train, validation, and test Sets

Table 1. Dataset statistics for training, validation, and testing subsets.

Number of messages	Training	Validation	Testing
Total	21,628	2,404	2,671
#Neutral	11,795	1,291	1,453
#Negative	5,491	637	683
#Positive	3,866	434	478
#Question	476	42	57
Avg. Words	27.21	27.18	27.12
Avg. Chars	89.82	89.50	90.36

Figures 2, 3, 4 and Table 1 clearly illustrate the imbalance among sentiment categories. The question class contains only 476 samples in the training set, 42 samples in the validation set, and 57 samples in the testing set. Classifying the question class is a very challenging task because it is very small. Many studies struggled with this class because of its very small size. The largest category in the dataset is the neutral class. It contains 11,795 samples in the training set, 1,291 samples in the validation set, and 1,453 samples in the testing set. The largest class size almost interferes with the smaller ones, such as the positive and negative classes, because models may learn patterns from neutral-class samples too much. In addition, the dataset contains highly informal, diverse language typical of Thai social media, including slang, abbreviations, emojis, mixed Thai-English content, and inconsistent spelling. These characteristics make the WISESIGHT dataset particularly challenging for sentiment analysis and require models to handle noisy and unpredictable real-world text. Therefore, the WISESIGHT dataset was utilized as a challenging benchmark to assess the performance of the proposed model.

### 3.2. Pre-Processing

The pre-processing stage is divided into two approaches. For the first approach, the input texts of the WISESIGHT dataset are directly tokenized using the WangchanBERTa tokenizer, and special tokens ([CLS], [SEP]) are added, where [CLS] represents the whole sentence and [SEP] marks the sentence boundary. The tokens are then converted into token IDs. These IDs are fed into the WangchanBERTa model. For example, the input sentence “สินค้าดีมาก จัดส่งรวดเร็ว” is tokenized into subword units using the WangchanBERTa tokenizer based on SentencePiece (e.g., [“สินค้า”, “ดี”, “มาก”, “จัดส่ง”, “รวดเร็ว”]). Special tokens [CLS] and [SEP] are added at the beginning and end of the sequence, respectively (e.g., [“[CLS]”, “สินค้า”, “ดี”, “มาก”, “จัดส่ง”, “รวดเร็ว”, “[SEP]”]). The resulting tokens are

then converted into token IDs (e.g., [101, 3456, 789, 1203, 4567, 8901, 102]), which are used as input to the WangchanBERTa model.

For the second approach, random oversampling was employed to address class imbalance in the WISESIGHT dataset by duplicating instances from minority classes. This approach helps mitigate model bias toward majority classes and improves the learning of underrepresented sentiments. Moreover, unlike synthetic data generation techniques, random oversampling preserves the original linguistic characteristics of the text. It is particularly suitable for Thai sentiment analysis. However, to ensure fair evaluation, experiments were also conducted on the original imbalanced dataset. After data balancing, training, validation, and testing sets for the balanced WISESIGHT dataset are shown in Table 3. The balanced WISESIGHT dataset is fetched to the next steps as the original one.

**Table 2.** The Balanced WISESIGHT dataset characteristics.

Number of messages	Training	Validation	Testing
Total	47,180	5,164	5,812
#Neutral	11,795	1,291	1,453
#Negative	11,795	1,291	1,453
#Positive	11,795	1,291	1,453
#Question	11,795	1,291	1,453

### 3.3. WangchanBERTa

Currently, contextual embeddings have been shown to outperform traditional word embeddings (e.g., Word2Vec or GloVe) in text representation. Unlike traditional word embedding, contextual embeddings capture the meaning of each word based on its surrounding context, allowing the same word to have different representations depending on its usage in a sentence. Therefore, contextual embeddings provide richer semantic representations and significantly improve performance in complex tasks such as sentiment analysis, particularly for languages like Thai with ambiguous word boundaries. For Thai texts, several BERT-based models have been developed to generate Thai contextual embeddings, including Thai-specific models such as WangchanBERTa and ThaiBERT, as well as multilingual models like mBERT and XLM-R. Among these, WangchanBERTa demonstrates superior performance due to its training on large-scale Thai corpora and its ability to better capture the linguistic characteristics of Thai. Therefore, WangchanBERTa is used to efficiently contextualize embeddings and is thus suitable for short, noisy, or informal Thai text, where fine-grained expressions drive sentiment.

In our work, the token IDs obtained from the pre-processing step are used as input to the WangchanBERTa model to generate contextual representations. First, each token ID is transformed into a dense vector through the embedding layer, which consists of token embeddings, segment embeddings, and positional embeddings. These components are combined to form the initial input representation, as defined in (1), where  $E_{\text{token}}$ ,  $E_{\text{segment}}$ , and  $E_{\text{position}}$  denote the word-piece, segment, and positional embeddings, respectively. The resulting embeddings are then passed through a stack of transformer encoder layers. Each layer consists of a self-attention mechanism and a position-wise feed-forward network. The self-attention mechanism, defined in (2), enables each token to attend to all other tokens in the sequence, where  $Q$ ,  $K$ , and  $V$  are the query, key, and value matrices derived from the input representation, and  $d_k$  denotes the dimension of the key vectors. The output of the attention layer is further refined by a feed-forward network, as defined in (3), where  $W_1$ ,  $W_2$ ,  $b_1$ , and  $b_2$  are learnable parameters. After passing through all encoder layers, the model produces contextualized representations  $X$ , as defined in (4), where each token representation  $x_t$  captures both semantic and contextual information,  $n$  is the sequence length, and  $d$  is the hidden dimension. These contextualized embeddings are then used as input features for downstream models, including CNN, BiLSTM, and BiGRU.

$$H^{(0)} = E_{\text{token}} + E_{\text{segment}} + E_{\text{position}} \quad (1)$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

$$\text{FFN}(H) = \text{GELU}(HW_1 + b_1)W_2 + b_2 \quad (3)$$

$$H = [h_1, h_2, \dots, h_n] \in \mathbb{R}^{n \times d} \quad (4)$$

### 3.4. Modeling

Traditional deep learning models such as BiLSTM, BiGRU, and CNN have been widely used for sentiment analysis due to their ability to capture sequential dependencies and local n-gram features. However, they rely on static embeddings and struggle to model rich contextual semantics, particularly in short, noisy, and linguistically complex Thai social media texts. CNN models further lack the ability to capture long-range dependencies, while recurrent models may be insufficient for handling highly ambiguous sentiment patterns (e.g., neutral vs. question) [32]. On the other hand, transformer-based models such as WangchanBERTa provide powerful contextualized representations and significantly improve Thai NLP performance, but they primarily focus on global context and may overlook fine-grained local features critical for sentiment classification in short texts [2,12]. Therefore, relying on either traditional deep learning or transformer models alone is insufficient. Therefore, integrating both approaches can effectively capture complementary global and local features.

We hybridize WangchanBERTa with BiLSTM, BiGRU, and CNN to mitigate noise in the WISIGHT dataset and improve classification performance while maintaining a relatively small parameter size. The proposed hybrid models are considered lightweight, as they incorporate only a single additional neural layer (BiLSTM, BiGRU, or CNN) on top of the WangchanBERTa model. This design avoids deep or complex stacking to reduce the number of trainable parameters and computational overhead while preserving strong performance. All three hybrid models share WangchanBERTa as the backbone, a 12-layer model with 768 hidden dimensions and 12 attention heads [12], serving as an initial contextual encoder to produce high-level hidden representations of Thai text as contextual embeddings.

The proposed WangchanBERTa-CNN model integrates WangchanBERTa with a CNN for sentiment classification. Let  $x_t \in \mathbb{R}^d$  denote the contextual embedding at time step  $t$  produced by WangchanBERTa. These contextual embeddings are then processed by one-dimensional convolutional layers with kernel sizes  $k \in \{2, 3, 4\}$ . Each kernel size extracts local n-gram features from the contextual embeddings [32–34]. Each convolution operation is performed as shown in (5), where  $*$  denotes the convolution operator, and  $W^{(k)}$  and  $b^{(k)}$  represent the weight and bias of the filter with kernel size  $k$ . The resulting feature maps are then subjected to a max-over-time pooling operation, as defined in (6), in order to extract the most salient features. The pooled features from different kernel sizes are concatenated to form a unified feature vector, as shown in (7), where  $\oplus$  denotes concatenation. Finally, the resulting feature vector is passed through a fully connected layer to produce the output logits, as defined in (8).

$$C^{(k)} = \text{ReLU}(W^{(k)} * H + b^{(k)}) \quad (5)$$

$$\hat{c}^{(k)} = \max_t C_t^{(k)} \quad (6)$$

$$\mathbf{c} = [\hat{c}^{(2)} \oplus \hat{c}^{(3)} \oplus \hat{c}^{(4)}] \quad (7)$$

$$\mathbf{z} = W_y \mathbf{c} + b_y \quad (8)$$

The proposed WangchanBERTa-BiLSTM model integrates WangchanBERTa with a BiLSTM network to capture both contextual and sequential dependencies in Thai text. Let  $x_t \in \mathbb{R}^d$  denote the contextual embedding at time step  $t$  produced by WangchanBERTa. The contextual embeddings are passed to a BiLSTM layer, which processes the sequence in both forward and backward directions, as shown in (15) and (16), respectively. Each LSTM unit within the BiLSTM architecture comprises a forget gate (9), input gate (10), candidate cell state (11), cell state (12), output gate (13), and hidden state (14), where  $W$  is the weight matrix applied to the contextual embedding  $x_t$ ,  $U$  is the weight matrix applied to the previous hidden state  $h_{t-1}$ , and  $b$  is the bias vector. These components collectively facilitate the learning of long-range dependencies and rich contextual representations. The hidden states from both directions are concatenated to form the final representation at each time step, as shown in (17). To obtain a fixed-length representation, the concatenated hidden state is selected at the final time step  $T$ , where  $T$  denotes the length of the sequence. Finally, the feature vector is passed through a fully connected layer, as shown in (18), where  $z$  denotes the output logits for the target sentiment classes.

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (9)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (10)$$

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (11)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (12)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (13)$$

$$h_t = o_t \odot \tanh(c_t) \quad (14)$$

$$\vec{h}_t = \text{LSTM}_f(x_t, \vec{h}_{t-1}) \quad (15)$$

$$\overleftarrow{h}_t = \text{LSTM}_b(x_t, \overleftarrow{h}_{t+1}) \quad (16)$$

$$h_t^* = \left[ \vec{h}_t \oplus \overleftarrow{h}_t \right] \in \mathbb{R}^{2d} \quad (17)$$

$$z = W_y h_T^* + b_y \quad (18)$$

The proposed WangchanBERTa-BiGRU model integrates WangchanBERTa with a BiGRU network to capture both contextual and sequential features in Thai text. Let  $x_t \in \mathbb{R}^d$  denote the contextual embedding at time step  $t$  generated by WangchanBERTa. These embeddings are then fed into a BiGRU layer, which processes the sequence in both forward and backward directions, as shown in (23) and (24), respectively. Each GRU unit operates through gating mechanisms that regulate the flow of information. Compared to LSTM, GRU adopts a simplified architecture with two primary gates: the update gate and the reset gate, as shown in (19) and (20), respectively. Based on these gates, the candidate hidden state  $\tilde{h}_t$  is computed by combining the current input  $x_t$  with the reset-modulated previous hidden state, as defined in (21). The final hidden state  $h_t$  is then obtained by interpolating between the previous hidden state and the candidate hidden state using the update gate, as shown in (22). This mechanism enables GRU to effectively capture sequential dependencies while maintaining lower computational complexity than LSTM. The hidden states from both directions are concatenated at each time step, as

shown in (25). To obtain a fixed-length representation, the concatenated hidden state at the final time step  $T$  is selected, where  $T$  denotes the sequence length. Finally, the feature vector is passed through a fully connected layer to produce the output logits, as shown in (26).

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (19)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \quad (20)$$

$$\tilde{h}_t = \tanh(W_h x_t + U_h (r_t \odot h_{t-1}) + b_h) \quad (21)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (22)$$

$$\vec{h}_t = \text{GRU}_f(x_t, \vec{h}_{t-1}) \quad (23)$$

$$\overleftarrow{h}_t = \text{GRU}_b(x_t, \overleftarrow{h}_{t+1}) \quad (24)$$

$$h_t^* = [\vec{h}_t \oplus \overleftarrow{h}_t] \in \mathbb{R}^{2d} \quad (25)$$

$$z = W_y h_T^* + b_y \quad (26)$$

For all proposed models, the final logits  $\mathbf{z}$  are converted into class probabilities using the softmax function, as shown in (27). The models are trained using the categorical cross-entropy loss, as shown in (28), where  $C$  denotes the number of classes,  $y_i$  is the ground-truth label, and  $\hat{y}_i$  is the predicted probability for class  $i$ . To improve generalization and mitigate overfitting, a dropout layer is applied to the extracted feature representation before the final classification layer. In practice, the softmax operation is internally combined with the cross-entropy loss during training.

$$\hat{y}_i = \frac{\exp(z_i)}{\sum_{j=1}^C \exp(z_j)} \quad (27)$$

$$\mathcal{L} = - \sum_{i=1}^C y_i \log(\hat{y}_i) \quad (28)$$

### 3.5. Evaluation

To evaluate model performance across categories in the WISESIGHT dataset, we use three standard classification metrics. The metrics include accuracy as shown in (29), precision, recall, and F1-Score. For multi-class evaluation, precision, recall, and F1-Score are computed per class as shown in (30), (31), and (32), respectively. In addition, the macro-average scores are used to evaluate the average of precision, recall, and F1-score as shown in (33), (34), and (35), respectively. Macro-average F1-score is adopted as the primary evaluation metric because the WISESIGHT dataset is highly imbalanced. Unlike accuracy or micro-average F1-score, macro-average F1-score assigns equal importance to all classes, ensuring that performance on minority classes is not overlooked. This provides a more reliable and fair assessment of model effectiveness in real-world sentiment classification scenarios.

$$Acc = \frac{\sum_{c=1}^4 TP_c}{N} \quad (29)$$

$$P_c = \frac{TP_c}{TP_c + FP_c} \quad (30)$$

$$R_c = \frac{TP_c}{TP_c + FN_c} \quad (31)$$

$$F1_c = 2 \times \frac{P_c \times R_c}{P_c + R_c} \quad (32)$$

$$\text{Macro avg. } P = \frac{1}{4} \sum_{c=1}^4 P_c \quad (33)$$

$$\text{Macro avg. } R = \frac{1}{4} \sum_{c=1}^4 R_c \quad (34)$$

$$\text{Macro avg. } F1 = \frac{1}{4} \sum_{c=1}^4 F1_c \quad (35)$$

From equations (29) - (35),  $c$  denotes each sentiment category. There are four categories: positive, neutral, negative, or question.  $TP_c$  refers to the number of correctly predicted samples for class  $c$ . For example,  $TP$  of the positive class means the number of correctly predicted samples as positive.  $N$  represents the total number of predictions.  $FP_c$  indicates the number of samples that do not belong to class  $c$  but were classified as class  $c$ . For example,  $FP$  for the positive class is the number of samples predicted as positive but belonging to the neutral, negative, or question class.  $FN_c$  represents the number of samples that belong to class  $c$  but were predicted as other classes. For example,  $FN$  for the positive class is the number of samples that belong to the positive class but were predicted as neutral, negative, or question. In addition, confusion matrices are utilized to analyze classification performance, and the number of model parameters is evaluated to verify the lightweight nature of the proposed models.

## 4. Results and Discussion

In the experimental setup, the input sequence length was limited to 128 tokens. The model was trained with a batch size of 16 and optimized using AdamW with a learning rate of  $2e^{-5}$ , incorporating a weight decay of 0.01 to mitigate overfitting. A dropout rate of 0.2 was also applied to further improve generalization. Together, weight decay and dropout serve as key regularization techniques in the training process. All our models were trained for 20 epochs to identify the optimal training point by monitoring the smallest gap between the training and validation losses. The optimal point of each model was used for classification.

### 4.1. Performance on the WISESIGHT Dataset

First, the WISESIGHT benchmark (without data balancing) is used in the experiment to compare our proposed models with the state-of-the-art (the parallel hybrid model [13]), and BiLSTM-CNN [1]. The results are shown in Table 3.

**Table 3.** The comparative performance on the WISESIGHT dataset (%)

Model	Acc	Negative			Neutral			Positive			Question			Macro			
		P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1	
BiLSTM-CNN [1]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	55.21
Parallel Hybrid [13]	73.64	76.62	76.85	76.74	74.53	<b>83.49</b>	78.76	65.53	43.56	52.33	51.46	36.87	42.96	67.04	60.19	62.70	
WangchanBERTa-BiLSTM	73.12	76.21	<b>83.02</b>	79.47	<b>77.71</b>	76.53	77.12	57.28	<b>51.05</b>	53.98	42.84	52.63	47.24	63.51	<b>65.81</b>	64.45	
WangchanBERTa-BiGRU	72.00	76.91	78.04	77.47	75.70	77.84	76.76	55.77	48.54	51.90	39.71	47.37	43.20	62.02	62.95	62.33	
WangchanBERTa-CNN [1,2,3]	<b>74.77</b>	<b>79.44</b>	78.62	79.03	76.28	82.30	<b>79.13</b>	62.47	50.84	<b>56.06</b>	<b>55.26</b>	36.84	44.21	<b>68.36</b>	62.15	64.62	
WangchanBERTa-CNN [2,3,4]	74.73	76.79	<b>83.89</b>	79.64	76.75	81.35	78.98	<b>65.94</b>	44.56	53.18	53.85	49.12	<b>51.38</b>	68.08	64.73	<b>65.80</b>	
WangchanBERTa-CNN [3,4,5]	73.79	77.19	81.21	79.17	77.06	79.08	78.06	62.67	49.16	55.10	37.21	<b>56.14</b>	44.76	63.53	<b>66.41</b>	64.27	
WangchanBERTa-CNN [1,2,3,4]	73.38	78.99	78.17	78.59	75.35	81.42	78.27	58.58	46.44	51.81	45.65	36.84	40.78	64.64	60.72	62.36	

\* N/A indicates that the corresponding metric was not reported in the original study

For the first experiment, we found that WangchanBERTa-CNN with kernel sizes [2, 3, 4] yields higher accuracy than the parallel hybrid models WangchanBERTa-BiLSTM and WangchanBERTa-BiGRU. Therefore, we further investigated the impact of convolutional kernel sizes on model performance. Multiple WangchanBERTa-CNN configurations were evaluated using different kernel combinations, including [1,2,3], [2,3,4], [3,4,5], and [1,2,3,4]. Kernel size combinations such as [1,2] or single kernel sizes were not selected because they provide limited coverage of textual patterns. Using only small kernels (e.g., [1,2]) restricts the model to very short n-gram features, which may fail to capture meaningful phrase-level context. Similarly, using a single kernel size limits feature diversity, as the model can only learn patterns within a fixed receptive field [32]. Since the kernel size combinations [1,2,3] and [2,3,4] achieved high macro-average F1-scores. Therefore, we further extended this idea by evaluating the configuration [1,2,3,4], which integrates both smaller and slightly larger receptive fields. The results of kernel size configurations are also shown in Table 2. Overall, the WangchanBERTa-CNN model with kernel sizes [2,3,4] achieves the highest macro-average F1-score at 65.80%, outperforming both smaller [1,2,3], larger [3,4,5], and [1,2,3,4] kernel settings as well as WangchanBERTa-BiLSTM and WangchanBERTa-BiGRU. Its macro-average F1-Score is also slightly higher than that of all competing models. It outperforms the state-of-the-art parallel hybrid model [13] by 4.94%. Our findings suggest that medium-sized kernels capture local linguistic patterns in Thai social media text more effectively than either narrower or broader receptive fields.

Across sentiment categories, the negative class consistently shows the strongest performance for all models. The WangchanBERTa-CNN [2,3,4] model achieves the highest recall (83.89%) and F1-score (79.64%) for this class. The model is particularly effective in identifying negative sentiment, which often contains clearer lexical or emotional cues. The neutral class follows with stable performance across models, although the parallel hybrid model has the highest recall (83.49%). In contrast, performance on the positive and question classes remains comparatively low across all models. For the positive class, the best F1-score is achieved by the WangchanBERTa-CNN [2,3,4] model (53.18%), but it still outperforms the state-of-the-art model. The WangchanBERTa-CNN [1,2,3] model achieves the highest F1-score for the positive class. The question class is the most challenging category due to its very small size. The WangchanBERTa-CNN [2,3,4] model achieves the highest F1-score of 51.38%, and its overall performance remains high. However, it still suffers from limited training data and the inherent ambiguity of interrogative structures. These results reinforce the dataset's imbalance and the linguistic complexity of Thai questions in social media contexts.

Figures 5–10 compare the confusion matrices of all our evaluated models and reveal clear differences in classification performance across sentiment categories. From the confusion matrices, the WangchanBERTa-CNN [2, 3, 4] model achieves the most accurate and balanced predictions. It is the best model for accurately identifying negative sentiment. Moreover, the WangchanBERTa-CNN model performs better than other configurations on the question class. The smaller [1,2,3] and larger [3,4,5] kernel sizes produce more misclassifications, particularly between neutral and positive texts. The results indicate that medium-sized convolutional kernels can more effectively capture key linguistic cues in short, informal Thai social media texts.

In summary, Table 2 and Figures 5–10 show that the combination of WangchanBERTa with CNN using medium kernel sizes [2, 3, 4] achieves the most robust performance across sentiment categories. Therefore, combining contextualized embeddings and local feature extraction is particularly effective for handling short, noisy, and linguistically diverse Thai social media texts.

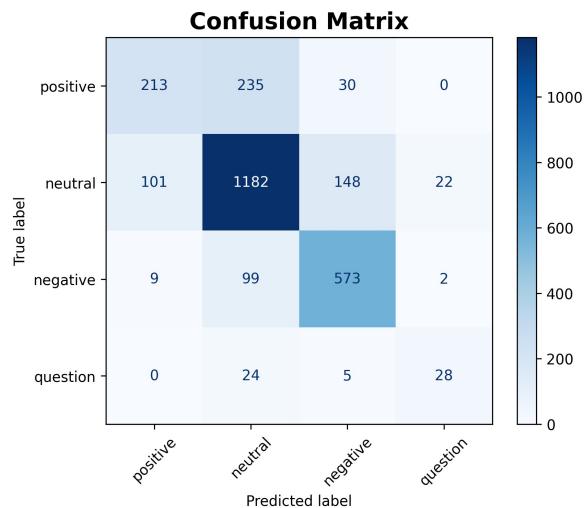


Figure 5. WangchanBERTa+CNN [2,3,4]

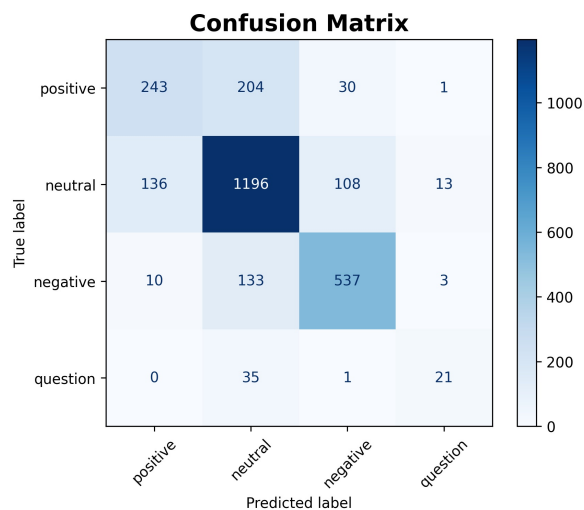


Figure 6. WangchanBERTa+CNN [1,2,3]

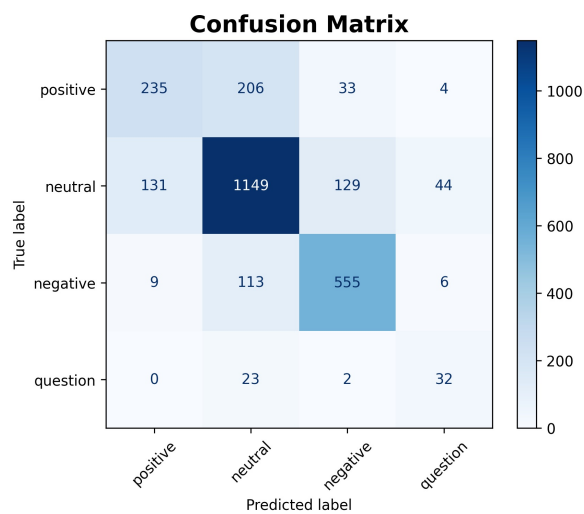


Figure 7. WangchanBERTa+CNN [3,4,5]

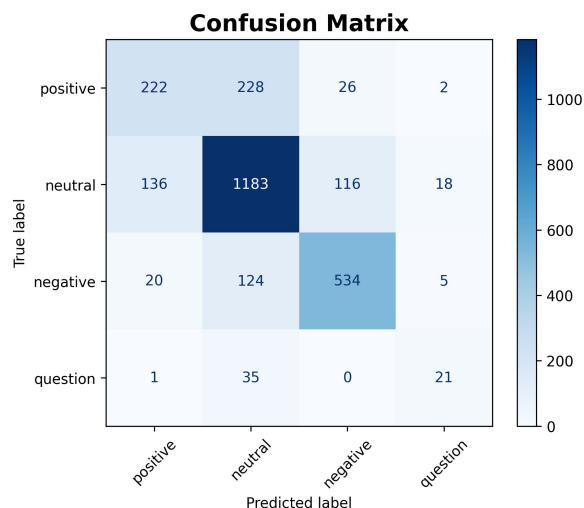


Figure 8. WangchanBERTa+CNN [1,2,3,4]

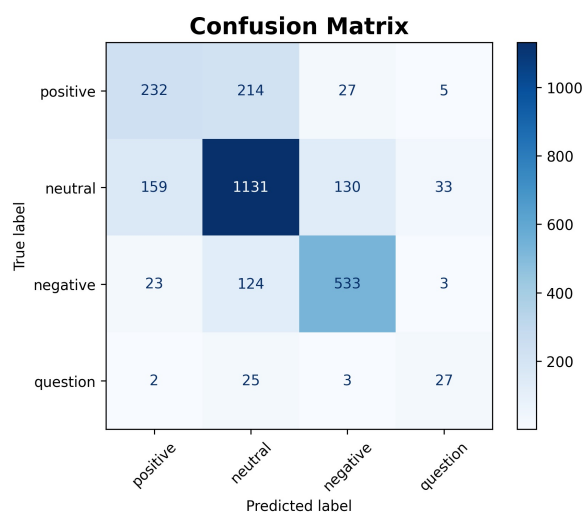


Figure 9. WangchanBERTa+BiGRU

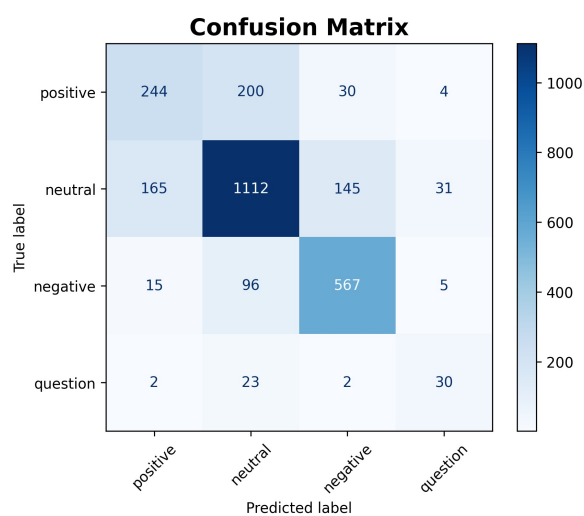


Figure 10. WangchanBERTa+BiLSTM

Next, we further enhanced the best-performing model, WangchanBERTa-CNN [2,3,4], by incorporating several advanced techniques, including attention mechanisms, squeeze-and-excitation (SE) blocks, gate fusion, and a three-model hybrid architecture. Attention-based approaches, including

single-head and multi-head attention, were applied to the concatenated CNN feature maps to dynamically assign importance weights to different time steps. A SE block is applied to each convolutional output. The SE block performs channel-wise recalibration by first applying global average pooling (squeeze operation) to capture channel-level statistics. For the gate fusion model, the parallel CNN branches with different kernel configurations ([1,2,3] and [2,3,4]) are utilized to capture multi-scale local patterns. To adaptively control feature importance, a gated fusion mechanism is applied. This mechanism enables the model to selectively emphasize or suppress features from different CNN branches. We also hybridize WangchanBERTa, BiLSTM, and CNN with kernel size [2, 3, 4] to enable complementary modeling of sequential dependencies and local n-gram features, while the attention mechanism further enhances the model by selectively focusing on the most informative contextual representations. Adding these techniques to WangchanBERTa-CNN [2,3,4] did not lead to performance improvements, as shown in Table 4. This may be due to several factors. First, the WangchanBERTa-CNN [2, 3, 4] model is already well-optimized, and additional components may introduce unnecessary complexity, leading to overfitting, especially given the limited and noisy nature of social media data. Second, techniques such as attention and fusion mechanisms may not effectively capture additional useful information beyond what the pretrained encoder already learns. Third, the class imbalance and subtle distinctions between sentiment categories (e.g., neutral vs. question) may limit the effectiveness of more complex architectures. Finally, increasing model complexity can make training more difficult and less stable, resulting in suboptimal generalization performance. In addition, adding some techniques to WangchanBERTa-CNN increases the number of parameters and makes the models more complex, as shown in Figure 11. Figure 11 illustrates the relationship between model performance and parameter size for only the proposed and extended models, as prior studies do not report parameter sizes. Figure 11 demonstrates that the proposed WangchanBERTa-CNN [2,3,4] model achieves the highest performance while maintaining a relatively lower parameter count compared to other models.

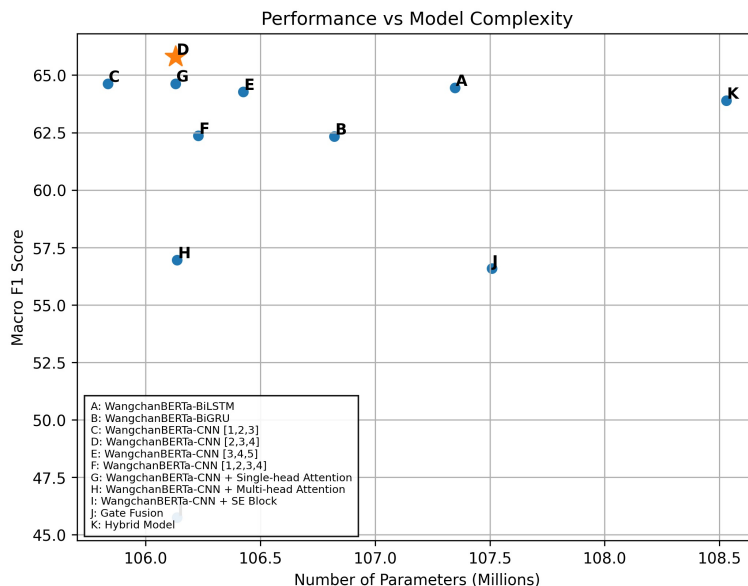


Figure 11. Performance vs complexity.

In summary, the results show that the simple combination of WangchanBERTa with CNN using medium kernel sizes, [2, 3, 4], achieves the most robust performance across sentiment categories while maintaining a relatively small parameter size. Therefore, combining contextualized embeddings and local feature extraction is particularly effective for handling short, noisy, and linguistically diverse Thai social media texts. The WangchanBERTa-CNN model achieved the 65.80% macro-average F1-score and improved the macro-average F1-score by approximately 4.94% when compared with the state-of-the-art, the parallel hybrid model [13].

**Table 4.** The comparative performance on the extended models (%)

Model	Acc	Negative			Neutral			Positive			Question			Macro		
		P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
WangchanBERTa-CNN+Single-head Attention	72.71	76.64	80.23	78.40	77.83	76.12	76.97	54.51	54.39	54.45	48.28	49.12	48.70	64.31	64.97	64.63
WangchanBERTa-CNN+Multi-head Attention	68.74	75.24	69.40	72.20	68.42	83.96	75.40	58.08	24.06	34.02	45.00	47.37	46.15	61.69	56.20	56.95
WangchanBERTa-CNN+SE Block	70.65	68.60	84.77	75.83	71.00	83.90	76.91	80.91	18.62	30.27	0.00	0.00	0.00	55.13	46.82	45.75
Gate Fusion	69.82	72.09	73.35	72.71	71.03	81.49	75.90	59.35	34.52	43.65	48.39	26.32	34.09	62.71	53.92	56.59
Hybrid Model	73.27	75.40	82.58	78.83	75.65	79.77	77.65	61.61	43.31	50.86	49.09	47.73	48.21	65.44	63.25	63.89

#### 4.2. Performance on the Balanced WISESIGHT Dataset

To ensure robust performance on the WISESIGHT dataset, models achieving a macro-average F1-score comparable to the best-performing model (i.e., greater than 64%) were selected for further analysis on the balanced WISESIGHT dataset. The WangchanBERTa-BiLSTM, WangchanBERTa-CNN [1,2,3], WangchanBERTa-CNN [2,3,4], and WangchanBERTa-CNN [3,4,5] models were experiment future and report the results as shown in Table 5.

The results in Table 5 clearly demonstrate that applying data balancing significantly improves model performance across all four architectures. Macro-average F1-score improvements exceeding 40% after balancing. The WangchanBERTa-CNN with kernel sizes [2,3,4] achieves the best overall performance. It achieve a macro-average F1-Score of 94.11%. The most pronounced improvements are observed in the "Question" and "Positive" classes. The balancing approach effectively addresses class imbalance and enhances minority-class recognition. Overall, these findings confirm that random oversampling plays a crucial role in boosting classification performance, especially when combined with an appropriately designed hybrid architecture.

Table 5. The improvement on the balanced WISESIGHT dataset (%)

Model	Acc	Negative			Neutral			Positive			Question			Macro		
		P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
<b>WangchanBERTa-BiLSTM (original)</b>	73.12	76.21	83.02	79.47	77.71	76.53	77.12	57.28	51.05	53.98	42.84	52.63	47.24	63.51	65.81	64.45
WangchanBERTa-BiLSTM (balancing)	93.52	94.16	95.32	94.74	92.55	83.77	87.94	89.02	95.39	92.10	98.50	99.59	99.04	93.56	93.52	93.45
Improvement (%)	+27.90	+23.55	+14.82	+19.21	+19.10	+9.46	+14.03	+55.41	+86.86	+70.62	+129.93	+89.23	+109.65	+47.32	+42.11	+45.00
<b>WangchanBERTa-CNN [1,2,3] (original)</b>	74.77	79.44	78.62	79.03	76.28	82.30	79.13	62.47	50.84	56.06	55.26	36.84	44.21	68.36	62.15	64.62
WangchanBERTa-CNN [1,2,3] (balancing)	93.82	93.47	95.53	94.49	92.33	84.39	88.18	91.38	95.46	93.37	98.11	100.00	99.05	93.82	93.84	93.77
Improvement (%)	+25.48	+17.66	+21.51	+19.56	+21.04	+2.54	+11.44	+46.28	+87.77	+66.55	+77.54	+171.44	+124.04	+37.24	+50.99	+45.11
<b>WangchanBERTa-CNN [2,3,4] (original)</b>	74.73	76.79	83.89	79.64	76.75	81.35	78.98	65.94	44.56	53.18	53.85	49.12	51.38	68.08	64.73	65.80
WangchanBERTa-CNN [2,3,4] (balancing)	94.14	94.50	94.50	94.50	89.91	88.24	89.07	93.68	93.81	93.75	98.31	100.00	99.15	94.10	94.14	<b>94.11</b>
Improvement (%)	+25.97	+23.06	+12.65	+18.66	+17.15	+8.47	+12.78	+42.07	+110.53	+76.29	+82.56	+103.58	+92.97	+38.22	+45.43	+43.02
<b>WangchanBERTa-CNN [3,4,5] (original)</b>	73.79	77.19	81.21	79.17	77.06	79.08	78.06	62.67	49.16	55.10	37.21	56.14	44.76	63.53	66.41	64.27
WangchanBERTa-CNN [3,4,5] (balancing)	93.84	93.47	95.53	94.49	92.33	84.39	88.18	91.38	95.46	93.37	98.11	100.00	99.05	93.82	93.84	93.84
Improvement (%)	+27.17	+21.09	+17.63	+19.35	+19.82	+6.71	+12.96	+45.81	+94.18	+69.46	+163.67	+78.13	+121.29	+47.68	+41.30	+46.01

## 5. Conclusions

This study presents lightweight hybrid models that integrate WangchanBERTa with deep learning architectures, including BiLSTM, BiGRU, and CNN, for Thai sentiment analysis on social media texts. The proposed approach effectively addresses key challenges in the WISESIGHT benchmark, including class imbalance, short text lengths, noisy user-generated content, and ambiguity between sentiment categories. Experimental results demonstrate that the WangchanBERTa–CNN model with optimal kernel size configuration ([2,3,4]) achieves the best performance on the WISESIGHT benchmark, with a macro-average F1-score of 65.80%, outperforming the state-of-the-art by 4.94% while maintaining a relatively low parameter size. Furthermore, the systematic analysis of CNN kernel sizes provides new insights into sentiment-specific feature extraction, highlighting the importance of combining contextualized embeddings with local feature learning. On the balanced WISESIGHT dataset, the WangchanBERTa–CNN model continues to achieve the best performance. In addition, dataset balancing significantly improves the performance of the proposed models, as random oversampling effectively enhances classification results. The WangchanBERTa–CNN model with kernel size configuration [2,3,4] achieves an excellent macro-average F1-Score of 94.11% on the balanced WISESIGHT dataset. These findings confirm that the proposed hybrid approach, the WangchanBERTa–CNN model with a kernel size configuration of [2,3,4], is effective for handling short, noisy, and linguistically diverse Thai social media texts.

While the final version of the proposed WangchanBERTa-CNN architecture with medium kernel sizes also shows strong performance, several avenues for future work remain. For instance, focal loss or class-balanced loss, SMOTE-style oversampling, and data augmentation strategies tailored to Thai social media can help address the under-representation of the positive and question classes. The incorporation of syntactic or semantic features, POS tags, dependency relations, and sentiment lexicons would improve the detection of more subtle sentiment signals, which CNNs alone fail to learn. If we extend this method to incorporate more general conversational context (threads, replies, or posts surrounding them), classification accuracy may improve, particularly for ambiguous sentiment. Testing the model on diverse Thai sentiment datasets or developing it in multi-domain settings (such as product reviews and political discourse) may enhance its robustness and generalization.

**Acknowledgments:** This research project was financially supported by Mahasarakham University.

## References

1. Pasupa, K.; Seneewong Na Ayutthaya, T. Hybrid Deep Learning Models for Thai Sentiment Analysis. *Cogn Comput*, **2022**, *14*, 167-193, doi: 10.1007/s12559-020-09770-0.
2. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers. *arXiv* **2018**, doi:10.48550/arXiv.1810.04805.
3. Wu, S.; Dredze, M. Beto, Bentz, Becas: The Surprising Cross-Lingual Effectiveness of BERT. *arXiv*, **2019**, doi: 10.48550/arXiv.1904.09077.
4. Zhuang, L.; Wayne, L.; Ya, S.; Jun, Z. A Robustly Optimized BERT Pre-training Approach with Post-training. In Proceedings of the 20th Chinese National Conference on Computational Linguistics, Huhhot, China, (2021), 471 - 484, doi: 10.1007/978-3-030-84186-7\_31.
5. Lan, Z.; Chen, M.; Goodman, S.; Gimpel, K.; Sharma, P.; Soricut, R. ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. *arXiv* **2019**, doi:10.48550/arXiv.1909.11942.
6. Conneau, A.; Khandelwal, K.; Goyal, N.; Chaudhary, V.; Wenzek, G.; Guzmán, F.; Grave, E.; Ott, M.; Zettlemoyer, L.; Stoyanov, V. Unsupervised Cross-lingual Representation Learning at Scale. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics Online, 8440-8451,(2020), doi: 10.18653/v1/2020.acl-main.747.
7. Clark, K.; Luong, M.-T.; Le, Q.; Manning, C. ELECTRA: Pre-training Text Encoders as Discriminators. *arXiv* **2020**, doi:10.48550/arXiv.2003.10555.
8. Gurgurov, D.; Bäuml, T.; Anikina, T. Multilingual Large Language Models and Curse of Multilinguality. *arXiv* **2024**, doi:10.48550/arXiv.2406.10602.

9. Chang, T.A.; Arnett, C.; Tu, Z.; Bergen, B.K. When Is Multilinguality a Curse? Language Modeling for 250 Languages. Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, 4074-4096, Miami, Florida, USA, (2024), doi:10.18653/v1/2024.emnlp-main.236.
10. Blevins, T.; Limisiewicz, T.; Gururangan, S.; Li, M.; Gonen, H.; Smith, N.A.; Zettlemoyer, L. Breaking the Curse of Multilinguality with Cross-lingual Expert Language Models. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, 10822-10837, Miami, Florida, USA, (2024), doi: 10.18653/v1/2024.emnlp-main.604
11. Zhao, Z.; Aletras, N. Comparing Explanation Faithfulness between Multilingual and Monolingual Fine-tuned Language Models. In Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), Mexico City, Mexico, (2024), 3226-3244, doi: 10.18653/v1/2024.naacl-long.178.
12. Lowphansirikul, L.; Polpanumas, C.; Jantrakulchai, N.; Nutanong, S. WangchanBERTa: Pretraining transformer-based Thai Language Models. *arXiv*, **2021**, doi: 10.48550/arXiv.2101.09635.
13. Suraratchai, K.; Phoomvuthisarn, S. Thai Language Sentiment Analysis with a Hybrid Method on WangchanBERTa-CNN-BiLSTM. *Journal of Information Science and Technology (JIST)*, **2024**, *14*, 1-11, Available online: <https://ph02.tci-thaijo.org/index.php/JIST/article/view/253881>.(accessed on 9 February 2026).
14. Phienthrakul, T.; Kijisirikul, B.; Takamura, H.; Okumura, M. Sentiment Classification with Support Vector Machines and Multiple Kernel Functions. In *Neural Information Processing. ICONIP 2009*; Leung, C.S., Lee, M., Chan, J.H., Eds.; Lecture Notes in Computer Science, vol 5864. Springer, Berlin, Heidelberg, 2009, 583-592, doi: 10.1007/978-3-642-10684-2\_65.
15. Lertsuksakda, R.; Netisopakul, P.; Pasupa, K. Thai sentiment terms construction using Hourglass of Emotions. 2014 6th International Conference on Knowledge and Smart Technology (KST), 46-50, Chon Buri, Thailand, (2014), doi: 10.1109/KST.2014.6775392.
16. Chirawichitchai, N. Emotion classification of Thai text using machine learning techniques. 2014 11th International Joint Conference on Computer Science and Software Engineering (JCSSE), 91-96, Chon Buri, Thailand, (2014), doi:10.1109/JCSSE.2014.6841848.
17. Chirawichitchai, N. Developing Term Weighting Scheme Based on Term Occurrence Ratio for Sentiment Analysis. In: Kim, K. (eds) *Information Science and Applications. Lecture Notes in Electrical Engineering*, vol 339. Springer, Berlin, Heidelberg, (2015), doi:10.1007/978-3-662-46578-3\_87.
18. Pasupa, K.; Netisopakul, P.; Lertsuksakda, R. Sentiment analysis of Thai children stories. *Artif Life Robotics*, **2016**, *21*, 357-364, doi: 10.1007/s10015-016-0283-8.
19. Netisopakul, P.; Pasupa, K.; Lertsuksakda, R. Hypothesis testing based on observation from Thai sentiment classification. *Artif Life Robotics*, **2017**, *22*, 184-190, doi: 10.1007/s10015-016-0341-2.
20. Haruechaiyasak, C.; Palingoon, P.; Trakultaweekoon, K. S-Sense: A Sentiment Analysis Framework for Social Media Monitoring. *Information Technology Journal KMUTNB* **2018**, *14*, 11-22.
21. Porntrakoon, P.; Moemeng, C. Thai Sentiment Analysis for Consumer's Review in Multiple Dimensions Using Sentiment Compensation Technique (SenseComp). In 2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 25-28, Chiang Rai, Thailand, (2018), doi: 10.1109/ECTICon.2018.8619892.
22. Tesmuang, R.; Chirawichitchai, N. Sentiment Analysis of Thai Online Product Reviews using Genetic Algorithms with Support Vector Machine. *Progress in Applied Science and Technology*, **2020**, *10*(2), 7-13, Available online: <https://ph02.tci-thaijo.org/index.php/past/article/view/241933>.(accessed on 5 February 2026).
23. Vateekul, P.; Koomsubha, T. A study of sentiment analysis using deep learning techniques on Thai Twitter data. In 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), Khon Kaen, Thailand, (2016), doi: 10.1109/JCSSE.2016.7748849.
24. Pasupa, K.; Seneewong Na Ayutthaya, T. Thai sentiment analysis with deep learning techniques: A comparative study based on word embedding, POS-tag, and sentic features. *Sustainable Cities and Society*, **2019**, *50*, 101615, doi: 10.1016/j.scs.2019.101615.
25. Thong-iad, K.; Netisopakul, P. Comparison of Thai Sentence Sentiment Tagging Methods Using Thai Sentiment Resource. In *Recent Advances in Information and Communication Technology 2019*; Boonyopakorn, P., Meesad, P., Sodsee, S., Unger, H. Eds.; Advances in Intelligent Systems and Computing, vol 936. Springer, Cham, 2019, 89-98, doi: 10.1007/978-3-030-19861-9\_9.
26. Jitboonyapinit, C.; Maneerat, P.; Chirawichitchai, N. Sentiment Analysis on Thai Social Media Using Convolutional Neural Networks and Long Short-Term Memory. *INTERNATIONAL SCIENTIFIC JOUR-*

- NAL OF ENGINEERING AND TECHNOLOGY (ISJET)*, **2023**, 7, 74-80, Available online: <https://ph02.tci-thaijo.org/index.php/isjet/article/view/246935>.(accessed on 9 February 2026).
27. Khamphakdee, N.; Seresangtakul, P. An Efficient Deep Learning for Thai Sentiment Analysis. *Data*, **2023**, 8, 90, doi: 10.3390/data8050090.
  28. Nokkaew, M.; Nongpong, K.; Yeophantong, T.; Ploykitikoon, P.; Arjharn, W.; Siritaratiwat, A.; Narkglom, S.; Wongsinlatam, W.; Remsungnen, T.; Namvong, A.; Surawanitkun, C. Analyzing online public opinion on Thailand-China high-speed train and Laos-China railway mega-projects using advanced machine learning for sentiment analysis. *Social Network Analysis and Mining*, **2023**, 14(1), 15, doi: 10.1007/s13278-023-01168-8.
  29. Satjathanakul, J.; Siriborvornratanakul, T. Sentiment analysis in product reviews in Thai language. *Int. J. Inf. Technol.*, **2025**, 17, 1979-1985, doi: 10.1007/s41870-024-01907-w.
  30. Emphan, C.; Tiamkaew, E.; Khruahong, S. Enhancing the Performance of Sentiment Analysis Models Using GridSearchCV: A Case Study on Electric Vehicles in Thailand. *Journal of Applied Informatics and Technology (JIT)* **2026**, 8, 260631. doi:10.14456/jait.2026.10.
  31. Suriyawongkul, A.; Chuangsuwanich, E.; Chormai, P.; Chantarapratin, N.; Prasertsom, P.; Sawatphol, J.; Yamada, N.; Rutherford, A.; Polpanumas, C.; Udomcharoenchaikit, C. PyThaiNLP/Wisesight Sentiment Corpus with Word Tokenization Label, *Zenodo*, **2024**, doi:10.5281/zenodo.3457446.
  32. Kim, Y. Convolutional Neural Networks for Sentence Classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1746–1751, Doha, Qatar, (2014).
  33. Minaee, S.; Kalchbrenner, N.; Cambria, E.; Nikzad, N.; Chenaghlu, M.; Gao, J. Deep Learning-based Text Classification: A Comprehensive Review. *ACM Computing Surveys (CSUR)* **2021**, 54, 1–40, doi: 10.1145/3439726.
  34. Kalchbrenner, N.; Grefenstette, E.; Blunsom, P. A Convolutional Neural Network for Modelling Sentences. *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference*, **2014**, doi: 10.3115/v1/P14-1062.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.