Investigating Hybrid GRU-Conv1D-BiGRU Architectures for Enhanced Sentiment Classification in Thai Crisis News Tweets

Sasiphan Nitayaprapha

ORCID: https://orcid.org/0009-0004-8361-0388
Faculty of Liberal Arts, Thammasat University

DOI: 10.29322/IJSRP.16.05.2026.p17314
https://dx.doi.org/10.29322/IJSRP.16.05.2026.p17314

Paper Received Date: 17th April 2026
Paper Acceptance Date: 13th May 2026
Paper Publication Date: 20th May 2026

Abstract- This research examines the effectiveness of hybrid deep learning architectures for sentiment classification in Thai-language social media, focusing on applications in crisis management. The study proposes and assesses eight configurations of a Hybrid Bidirectional Gated Recurrent Unit (HBiGRU) architecture, employing a specialized GRU-Conv1D-BiGRU pipeline. This hybrid architecture assists in capturing both localized n-gram features and global temporal dependencies in news tweets. The experimental outcomes were benchmarked against a well-established BiGRU baseline. Results reveal that while hybrid models exhibited competitive performance and unique advantages in identifying specific sentiment categories, the conventional BiGRU proved to be the most reliable tool for overall classification accuracy on this particular dataset. The study concludes that for linguistically rich and unsegmented languages like Thai, enhancing architectural complexity through convolutional filtering may introduce structural noise. Thus, a simplified bidirectional approach remains highly effective for real-time crisis management.

Index Terms- Thai Sentiment Analysis, Deep Learning in Crisis Management, HBiGRU, Convolutional Neural Networks (Conv1D), NLP for Unsegmented Languages

I. Introduction

Social media platforms have emerged as critical sensors for real-time information during public emergencies and societal crises [1]. In the context of Thailand, news tweets provide a high-frequency stream of data that is vital for government agencies and emergency responders It is helpful for these agencies and responders to gauge public sentiment from crisis-related news tweets and coordinate crisis management efforts effectively [2, 3]. However, the automated analysis of this data is significantly hindered by the inherent complexities of the Thai language [4, 5]. Unlike English, Thai is a non-segmented language characterized by a lack of word boundaries, a complex system of tonal markers, and the frequent use of informal slang in social media discourse [4, 5, 6].

The primary challenge lies in the linguistic density and ambiguity of news tweets during crises [2, 3]. Within these tweets, sentiment is often nuanced or buried within factual reporting [2, 7]. Traditional Recurrent Neural Network (RNN) architectures frequently struggle to capture both the long-range dependencies and the localized textual features simultaneously [8, 9]. Bidirectional Gated Recurrent Units (BiGRU) have shown promise in handling sequential dependencies by processing data in both forward and backward directions [10, 11].
However, there remains a critical need for more robust hybrid architectures capable of extracting multi-scale features from the noisy and unsegmented Thai text in crisis-related news [4, 8, 6, 12].

This research aims to address these limitations by investigating the efficacy of a hybrid architecture, specifically the Hybrid Bidirectional Gated Recurrent Unit (HBiGRU). The research integrates Conv1D layers for local feature extraction with BiGRU and GRU layers for global context. This study explores whether such hybrid models can significantly improve upon standard BiGRU baselines. The objective is to identify an optimal configuration that balances model complexity with predictive stability. The outcomes will enhance sentiment classification accuracy for Thai-language crisis management. leading journals to complete their grades. In addition, the published research work also provides a big weight-age to get admissions in reputed varsity. Now, here we enlist the proven steps to publish the research paper in a journal.

II. METHODS

A. Data Collection

The dataset utilized for this study consists of Thai-language news tweets that were specifically curated for crisis management scenarios. Data was collected from various prominent Thai news channels on the Twitter (X) platform. This dataset offers a diverse stream of professional news reporting. A primary focus within this dataset is centered on COVID-19 related content. This specific focus ensures that the research targets a period of sustained crisis that possesses a very high thematic density.

B. Baseline Selection

To measure the efficacy of the proposed hybrid approach, a standard BiGRU baseline was selected to serve as the established benchmark. This architecture incorporates a bidirectional layer with 256 units, followed by a dense layer with 128 units, a dropout layer, and a final dense layer with 3 units for classification. During previous experimentation, this specific baseline model reached an accuracy of 74.22%. This result represented the strongest performance among the eight counterpart BiGRU models with different architectures that were tested on this exact Thai dataset. By using this baseline, a standard is set to evaluate whether adding convolutional layers and preliminary gated layers can actually yield statistically significant improvements.

C. HBiGRU Architecture

The HBiGRU configurations are outlined in Figure 1.

Model 1

Layer (type)	Output Shape
embedding (Embedding)	(None, 1335, 1335)
gru (GRU)	(None, 1335, 64)
conv1d (Conv1D)	(None, 1330, 128)
bidirectional (Bidirectional)	(None, 256)
dense (Dense)	(None, 64)
dropout (Dropout)	(None, 64)
dense_1 (Dense)	(None, 64)
dropout_1 (Dropout)	(None, 64)
dense_2 (Dense)	(None, 3)

Model 2

Layer (type)	Output Shape
embedding (Embedding)	(None, 1335, 1335)
gru (GRU)	(None, 1335, 64)
conv1d (Conv1D)	(None, 1330, 128)
bidirectional (Bidirectional)	(None, 256)
dense (Dense)	(None, 64)
dropout (Dropout)	(None, 64)
dense_1 (Dense)	(None, 3)

Model 3

Layer (type)	Output Shape
embedding (Embedding)	(None, 1335, 1335)
gru (GRU)	(None, 1335, 128)
conv1d (Conv1D)	(None, 1330, 128)
bidirectional (Bidirectional)	(None, 256)
dense (Dense)	(None, 64)
dropout (Dropout)	(None, 64)
dense_1 (Dense)	(None, 64)
dropout_1 (Dropout)	(None, 64)
dense_2 (Dense)	(None, 3)

Model 4

Layer (type)	Output Shape
embedding (Embedding)	(None, 1335, 1335)
gru (GRU)	(None, 1335, 64)
conv1d (Conv1D)	(None, 1330, 128)
bidirectional (Bidirectional)	(None, 256)
dense (Dense)	(None, 128)
dropout (Dropout)	(None, 128)
dense_1 (Dense)	(None, 3)

Model 5

Layer (type)	Output Shape
embedding (Embedding)	(None, 1335, 1335)
gru (GRU)	(None, 1335, 64)
conv1d (Conv1D)	(None, 1330, 128)
bidirectional (Bidirectional)	(None, 256)
dense (Dense)	(None, 128)
dropout (Dropout)	(None, 128)
dense_1 (Dense)	(None, 3)

Model 6

Layer (type)	Output Shape
embedding (Embedding)	(None, 1335, 1335)
gru (GRU)	(None, 1335, 128)
conv1d (Conv1D)	(None, 1330, 128)
bidirectional (Bidirectional)	(None, 256)
dense (Dense)	(None, 128)
dropout (Dropout)	(None, 128)
dense_1 (Dense)	(None, 3)

Model 7

Layer (type)	Output Shape
embedding (Embedding)	(None, 1335, 1335)
gru (GRU)	(None, 1335, 32)
conv1d (Conv1D)	(None, 1330, 128)
bidirectional (Bidirectional)	(None, 128)
dense (Dense)	(None, 64)
dropout (Dropout)	(None, 64)
dense_1 (Dense)	(None, 3)

Model 8

Layer (type)	Output Shape
embedding (Embedding)	(None, 1335, 1335)
gru (GRU)	(None, 1335, 64)
conv1d (Conv1D)	(None, 1330, 128)
bidirectional (Bidirectional)	(None, 256)
dense (Dense)	(None, 128)
dropout (Dropout)	(None, 128)
dense_1 (Dense)	(None, 64)
dropout_1 (Dropout)	(None, 64)
dense_2 (Dense)	(None, 3)

Figure 1: Layer Architecture of the HBiGRU Models

The core of the methodology involves the design and systematic testing of eight distinct Hybrid Bidirectional Gated Recurrent Unit (HBiGRU) configurations. As illustrated in Figure 1, each model utilizes a hierarchical structure designed to process the 1,335-length input sequences through three primary phases:

- Preliminary Gated Layer: An initial GRU layer (varying between 32, 64, or 128 units) acts as a temporal encoder to capture basic sequential patterns.
- Convolutional Feature Extraction: A Conv1D layer with 128 filters follows the first GRU, extracting localized n-gram features and reducing spatial dimensionality.
- Bidirectional Contextualization: A final Bidirectional layer (using either 256 or 128 units) processes the filtered features in both forward and backward directions to consolidate global context.

The eight specific models shown in Figure 1 were differentiated by changing the unit counts within the GRU, BiGRU, and dense layers. Additionally, the depth of the classification head was adjusted to range from one to two hidden dense layers, complete with associated dropout. This systematic variation allows for the identification of the best balance between the model's capacity for feature extraction and the actual ability to generalize across a dataset of Thai news tweets.

III. RESULTS

A. Comparative Performance Analysis

The performance metrics for the eight HBiGRU configurations and the baseline BiGRU model are summarized in Table 1. Based on the experimental results, Model 5 and Model 7 emerged as the top-performing hybrid architectures, achieving accuracies of 0.7159 and 0.7146, respectively.

Table 1: Performance Comparison of HBiGRU Configurations

Model	Accuracy	Precision	Recall	f1-score
1	0.6821	0.6845	0.6821	0.6827
2	0.6946	0.7037	0.6946	0.6964
3	0.6295	0.6338	0.6295	0.6312
4	0.6834	0.7049	0.6834	0.6830
5	0.7159	0.7188	0.7159	0.7132
6	0.6421	0.6379	0.6421	0.6345
7	0.7146	0.7259	0.7146	0.7150
8	0.6633	0.6689	0.6633	0.6644
BiGRU	0.7334	0.7348	0.7334	0.7324

B. Training Dynamics and Convergence

Figures 2 and 3 illustrate the training trajectories for the two leading hybrid models. Both models exhibit a characteristic learning pattern for Thai-language sentiment analysis:

• Model 5 Analysis: As shown in Figure 2, Model 5 reaches peak validation accuracy within 10 epochs. However, a divergence between training and validation loss emerges after the 5th epoch. This suggests that while the model effectively captures the linguistic nuances of Thai news tweets, it remains susceptible to the high variance inherent in social media data.

• Model 7 Analysis: Figure 3 highlights the efficiency of a more lightweight configuration. Despite having the smallest GRU layer (32 units) among all tested configurations, Model 7 achieves the highest F1-score of 0.7150. This suggests that for crisis management news, a less complex recurrent layer may actually improve generalization by reducing overfitting to specific tweet structures.

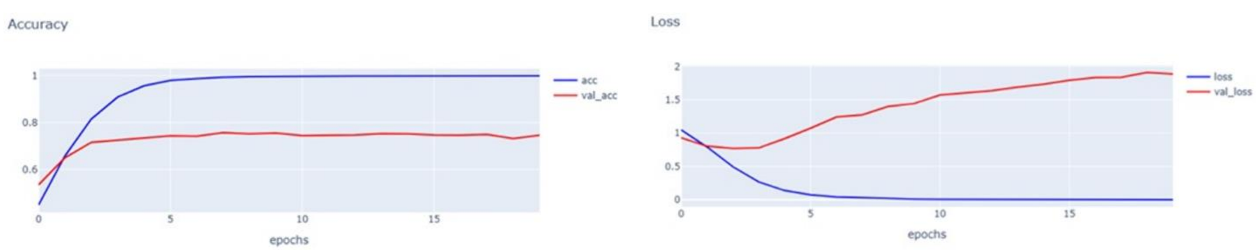

Figure 2: Accuracy and Loss Curves for the Optimized HBiGRU Model 5

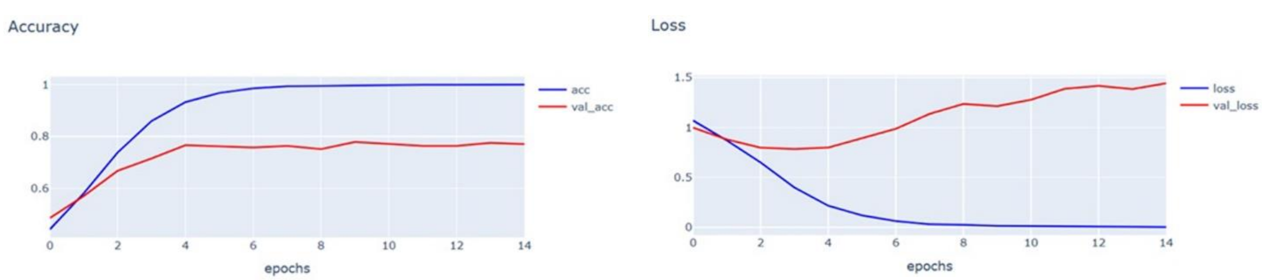

Figure 3: Accuracy and Loss Curves for the HBiGRU Model 7

C. Discussion of Architectural Impact

The results indicate that the GRU-Conv1D-BiGRU sequence successfully extracts multi-scale features from Thai text. Interestingly, while the baseline BiGRU maintains a slight lead in overall accuracy (0.7334), the hybrid models—specifically Model 7—demonstrate competitive precision (0.7259). Such precision is critical in crisis management to avoid false alarms in sentiment detection. The stability of the accuracy curves across these models confirms that the hybrid approach provides a consistent framework for processing the complexities of Thai-language Twitter data.

D. Class-Specific Performance Analysis

To further evaluate the robustness of the HBiGRU configurations, a class-specific analysis was conducted across Neutral, Positive, and Negative sentiments, as detailed in Table 2.

Insights from Class-Specific Data:
Superiority in Negative Sentiment Detection: All models, particularly the hybrid architectures, demonstrate their highest F1-scores in the Negative category. Both Model 5 and Model 7 achieved F1-scores of approximately 0.76. This performance indicates that the GRU-Conv1D-BiGRU pipeline is highly effective at identifying crisis-related warnings and distress in Thai tweets.

• Model 5 and Positive Sentiment: Interestingly, Model 5 significantly outperformed the baseline BiGRU in the Positive category, with an F1-score of 0.7623 compared to the baseline's 0.7089. This suggests that the larger dense layer configuration in Model 5 (128 units) better captures the specific linguistic patterns associated with optimistic news during a crisis.
• Model 7 and Neutral Sentiment: Model 7 achieved the highest Neutral F1-score (0.7052) among the hybrid models. Its high recall for neutral tweets (0.8025) suggests that a more streamlined GRU layer (32 units) is less likely to over-interpret factual news reporting as having a strong emotional bias. Such bias is a common challenge in Thai crisis management datasets.
• Recall vs. Precision Trade-offs: The data reveals that while the baseline BiGRU achieves the highest overall accuracy (0.7334), Model 7 provides a superior recall for the Neutral category (0.8025 compared to the baseline's 0.6296). Although the baseline maintains a stronger recall for Negative tweets (0.8129), Model 7 offers a more balanced performance for identifying Neutral information. This improved Neutral recall is critical to ensure that factual, objective reports are correctly categorized and not overlooked amidst high-emotion content in the context of Thai crisis management

Table 2: Breakdown of F1-scores by sentiment class for key configurations

Model	**Accuracy**	**Neutral**			**Positive**			**Negative**		
		precision	*recall*	*f1-score*	*precision*	*recall*	*f1-score*	*precision*	*recall*	*f1-score*
1	0.6821	0.6302	0.6872	0.6575	0.6543	0.6331	0.6435	0.7623	0.7266	0.7440
2	0.6946	0.6626	0.6626	0.6626	0.6322	0.7266	0.6767	0.8101	0.6906	0.7456
3	0.6295	0.6017	0.5844	0.5929	0.5418	0.5827	0.5615	0.7538	0.7158	0.7343
4	0.6834	0.5640	0.7613	0.6480	0.7665	0.5432	0.6358	0.7664	0.7554	0.7609
5	0.7159	0.6759	0.7037	0.6839	0.7556	0.6655	0.7623	0.8016	0.7266	0.7623
6	0.6421	0.6167	0.4568	0.5248	0.6041	0.6367	0.6200	0.6902	0.8094	0.7450
7	0.7146	0.6290	0.8025	0.7052	0.7292	0.6295	0.6757	0.8072	0.7230	0.7628
8	0.6633	0.7095	0.6132	0.6578	0.5756	0.6439	0.6078	0.7266	0.7266	0.7266
BiGRU	0.7334	0.7286	0.6296	0.6755	0.6765	0.7446	0.7089	0.7986	0.8129	0.8057

IV. DISCUSSION

A. Comparative Analysis

Hybrid vs. Standard Architectures
A critical finding of this study is that despite the architectural sophistication of the HBiGRU models, the baseline BiGRU model maintained a performance lead with an accuracy of 0.7334. The integration of Conv1D layers in the hybrid configurations was intended to enhance local feature extraction. However, the results suggest that for this specific Thai crisis dataset, the additional layers may have introduced structural noise. While Model 5 and Model 7 showed strong competitiveness—particularly in specific class F1-scores—the standard BiGRU's ability to maintain a superior global context without the pooling or filtering effects of a convolutional layer proved more effective for overall accuracy.

B. Thai Linguistic Factors and Model Interaction

The performance gap can be partially attributed to the unique linguistic characteristics of the Thai language.

Sequential Dependency
Thai is a non-segmented language where meaning is heavily dependent on long-range context rather than isolated tokens. The standard BiGRU architecture excels at capturing these bidirectional dependencies across long sequences (1,335 length) without the localized window constraints imposed by a Conv1D layer.

Tonal and Semantic Nuance
Sentiment in Thai news tweets is often conveyed through subtle tonal markers and particles located at the end of sentences. A pure BiGRU approach treats the entire sequence with equal temporal

importance, whereas the GRU-Conv1D-BiGRU pipeline may inadvertently down-sample these critical late-sequence markers during the convolutional phase.

Informal Social Media Structure

The frequent use of slang and irregular word boundaries in tweets complicates the fixed-filter approach of Conv1D. The BiGRU's gated mechanism is naturally more robust at ignoring noisy informalities to focus on the broader sentiment.

C. Computational Efficiency and Trade-offs

The study reveals a clear trade-off between architectural complexity and actual classification gains.

• Parameter Efficiency: Model 7, the most lightweight hybrid configuration (32 GRU units), achieved an accuracy of 0.7146, nearly matching the more complex Model 5. This suggests that increasing the width of the model—such as the 128 units used in Model 6—does not yield proportional gains in Thai sentiment classification and may lead to diminishing returns.

• Practical Application: For real-time crisis management, the marginal accuracy difference between the baseline and the hybrid models must be weighed against computational overhead. The stability of the standard BiGRU suggests it remains a highly efficient architecture for Thai NLP.

V. CONCLUSION

A. Summary of Findings

The experimental evaluation of eight Hybrid Bidirectional Gated Recurrent Unit (HBiGRU) configurations against a BiGRU baseline reveals that architectural complexity does not always translate to superior performance in Thai-language sentiment analysis. For this specific dataset of crisis-related news tweets, the BiGRU baseline model remains the most effective tool, achieving a peak accuracy of 0.7334. While the hybrid GRU-Conv1D-BiGRU approach showed notable strengths—specifically Model 5 in positive sentiment detection and Model 7 in neutral recall—the standard BiGRU's ability to preserve the full bidirectional context of unsegmented Thai text proved more robust for overall sentiment classification. These results emphasize that for linguistically complex languages like Thai, a deeper model is not necessarily a more accurate one.

B. Future Work

While the current hybrid designs did not surpass the baseline, they provide a valuable foundation for future architectural tuning. Future research should explore the following directions:

• Attention Mechanisms: Integrating Self-Attention or Multi-Head Attention layers could allow the model to weight specific sentiment-bearing Thai particles more effectively than fixed-size convolutional filters.

• Alternative Hybrid Sequences: Testing different layer orderings, such as placing the Conv1D layer before the initial GRU, might better facilitate early-stage feature extraction from raw, unsegmented character sequences.

• Pre-trained Language Models: Combining the current HBiGRU architecture with pre-trained Thai embeddings, such as WangchanBERTa, could provide the models with a stronger semantic baseline before fine-tuning on crisis-specific news.

• Hyperparameter Optimization: Further investigation into varying the kernel sizes of the Conv1D layers or implementing dynamic dropout rates could help mitigate the overfitting observed in the more complex HBiGRU variants.

REFERENCES

[1] M. Imran, C. Castillo, F. Diaz, and J. Vieweg, "Processing Social Media Messages in Mass Emergency: A Survey," ACM Computing Surveys, 2015, 47, 4, Article 67, pp. 1–38.

[2] S. Shayaa, et al., "Sentiment Analysis of Big Data: Methods, Applications, and Open Challenges," IEEE Access, 2018, 6, pp. 37807–37827.

[3] K. Matarat, "Aspect-Based Sentiment Analysis in Thai Texts: A Comparative Study of Machine Learning and Neural Network Approaches," Indochina Applied Sciences, 2025, 14, 3, https://doi.org/10.55674/ias.v14i3.262648.

[4] L. Lowphansirikul, C. Polpanumas, N. Jantrakulchai, and S. Nutanong, "WangchanBERTa: Pretraining transformer-based Thai Language Models," arXiv preprint arXiv:2101.09635, 2021.

[5] V. Sornlertlamvanich, "Word segmentation for Thai in machine translation system," Machine Translation, 1993, pp. 50–56.

[6] S. Otter, J. Medina, and J. Kalita, "A Survey of the Usages of Deep Learning for Natural Language Processing," IEEE Transactions on Neural Networks and Learning Systems, 2021, 32, 2, pp. 604–624.

[7] Y. Wang, M. Huang, and X. Zhu, "Attention-based LSTM for Aspect-level Sentiment Classification," Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2016, pp. 606–615.

[8] S. Minaee et al., "Deep Learning--based Text Classification: A Comprehensive Review," ACM Computing Surveys, 2021, 54, 3, pp. 1–40.

[9] K. Cho et al., "On the Properties of Neural Machine Translation: Encoder–Decoder Approaches," Proceedings of SSST-8, 2014, pp. 103–111.

[10] S. O. Alhumoud, "CNN-BiGRU-Focus: A Hybrid Deep Learning Classifier for Sentiment and Hate Speech Analysis of Ashura-Arabic Content for Policy Makers," International Journal of Advanced Computer Science and Applications, 2024, 15, 11, pp. 1006–1020.

[11] M. Schuster and K. K. Paliwal, "Bidirectional Recurrent Neural Networks," IEEE Transactions on Signal Processing, 1997, 45, 11, pp. 2673–2681.

[12] H. Jelodar et al., "Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a survey," Multimedia Tools and Applications, 2019, 78, pp. 15169–15211.

AUTHORS

Asst.Prof.Dr.Sasiphan Nitayaprapha
Ph.D. in Informatics (The University of Manchester)
Faculty of Liberal Arts, Thammasat University,
contact: sasiphan@tu.ac.th