*ICIC Express Letters* ICIC International ©2023 ISSN 1881-803X
Volume **17**, Number **1**, January **2023** pp. **91–101**

# THAI TEXT CLASSIFICATION EXPERIMENT USING CNN AND TRANSFORMER MODELS FOR TIMELY-TIMELESS CONTENT MARKETING

Pagon Gatchalee, Sajjaporn Waijanya* and Nuttachot Promrit*

Center of Excellence in AI and NLP
Department of Computing
Faculty of Science
Silpakorn University
Sanam Chandra Palace Campus, Rajamunka-nai Rd., Muang District
Nakhon Pathom 73000, Thailand
gatchalee_p@su.ac.th; *Corresponding authors: { waijanya_s; promrit_n }@silpakorn.edu

Received April 2022; accepted June 2022

Abstract. *The "timely" and "timeless" concepts are essential content marketing strategies currently applied in the industry. Timely is applied for the purpose of making viral content, whereas timeless is adopted to gain brand awareness and marketing performance in the long term for products that are not just meant to be temporal trends. This research aims to study and find the best text classification model for timely-timeless content classification in the Thai context. There are six timely-timeless text classification models from two major state-of-the-art methods which are sub-divided as follows: 1) Convolutional Neural Network (CNN) and its variants: CNN Conv1D, CNN Conv1D Skip-gram, and CNN Conv1D with fastText Skip-gram, and 2) Transformer and its variants: BERT, RoBERTa, and WangchanBERTa. Results reveal that, among these, WangchanBERTa, the large pre-trained corpus exclusively for the Thai language, has the best performance. It also has the highest validation accuracy at 93.06%, while testing accuracy is more than 90% at 93.00%. Apart from its accurate performance, the validation loss of WangchanBERTa is also the lowest, at only 23.85%. Furthermore, this transformer-based text classification model is considered the best, with performance metrics of 93% for both Precision and Recall and the highest F1-score at 92.00%. Although it is limited in the small dataset size of 600 articles with at least 250 words for each article, this dataset is separated into 336 training, 144 validation, and 120 test dataset articles. Therefore, this research can be a guideline for text classification on a small Thai-language dataset. The other contributions of this research are its proposal of performing preprocessing for cleaning Facebook datasets based on the timely and timeless classification purposes, and the introduction to the experiment of the details of model hyperparameter tuning, which can lead to high performance.*

**Keywords:** Timely, Timeless, Content marketing, Deep learning, CNN, Transformer, BERT, RoBERTa, WangchanBERTa

1. **Introduction.** One of the most common issues in the field of content marketing-based content creation is the inconsistency between the generated article and the style of the underlying or established theme. In particular, styles may not be the only form of content structuring. For example, the structure of news-style content generally contains four parts: 1) Headline, the first part which concludes the content as title and gains more attention from readers; 2) Lead, the introductory paragraph where crucial information is stated; 3) Body, the part of the content where important information is placed aside from the Headline and Lead parts; and 4) Tail, which is the summary where other information may be included [1]. However, it is also caused by the "timely" and "timeless" content

DOI: 10.24507/icicel.17.01.91

writing styles (also known as "ephemeral" and "evergreen" contents). These are article writing styles that are widely used today for content marketing, especially online content, such as websites, blogs, and social media [2-6].

The two main differences between timely and timeless contents are their time-sensitive and content objectives [7]. Timeless content is not time-sensitive, which means it is long-lived and not reliant on trends [8]. For example, the content of Chinese history, is considered "timeless". On the contrary, content regarding the COVID-19 situation in China is deemed "timely" due to the time-sensitive nature of such information [9]. Another difference is that the objective of timeless content is creating interesting content for audiences based on a marketing objective for a long-term marketing strategy rather than a short-term one. One example is creating the content for China's Tiananmen Square attraction guide for Facebook pages with China-related content or publishing in websites for long-term search engine optimization marketing purposes [10]. In contrast, in timely content, the primary objective is to create topical and trendy content, such as news and event updates, which are adopted in short-term marketing campaign strategies [11]. This style is used in creating traveling content by using buzzwords or words that became viral at a certain point in time. Viral content is also a typical objective of creating timely content [12].

Several research studies are on timely and timeless content classification, especially on English content, but written in other languages, it is relatively small, especially in Thai language. Nevertheless the popularity of text classification research has continued to increase [13]. Thus, it is challenging to investigate and find the best way to classify Thai-language timely and timeless content that benefit both content marketing and Thai text classification research.

Therefore, the current research aims to find the most efficient way to classify timeless and timely articles by applying existing algorithms categorized into two major state-of-the-art methods: Convolutional Neural Network (CNN) and the Transformer model. This work's contributions to the literature can be summarized as follows.

1) It conducts an experiment and finds the best classification model for Thai timely and timeless content by using Facebook page, "Aizhongchina", a Chinese affairs-related content provider and TOP5 of 2020 RAiNMaker Thailand Content creator awards as a case study.
2) It proposes a method to clean the Facebook page dataset prior to the timely and timeless classification process.
3) It presents a way of model tuning with hyperparameter adjustment and model-related factors to obtain the best performance for a given dataset.

The remainder of this paper is organized as follows: the related works are reviewed in Part 2, then the research process is illustrated in Part 3 of this paper, followed by the experiment results with the proposed method in Part 4, and a presentation of the conclusion and recommendations for future research in Part 5.

## 2. Related Works.

**2.1. Timely and timeless content.** StumbleUpon (Mix.com – social media platform) is one of the relatively well-known timely-timeless classification datasets used in the 2013 Kaggle competition. Some studies have also been published after that competition. In the work of web page classification as evergreen or ephemeral based on content [14], linear SVM gained an accuracy of around 87%. The authors first pre-processed the stop words and performed outlier removal. Another work [15], obtained a slightly higher accuracy rate of 88% by combination of a text preprocessing method, namely, the Term Frequency-Inverse Document Frequency (TF-IDF) technique, the singular value decomposition technique for noise reduction, along with non-textual features, and logistic regression to classify ephemeral and evergreen contents. Apart from English content, it only represents a quite small timely-timeless content classification research.

2.2. **Text classification.** Text classification is a natural language processing problem that has received much attention. It aims to divide messages into groups or wanted types. Messages can be classified from short messages with just a few sentences to many messages in the form of documents or articles. Research on long message segmentation or document classification has become more popular, with examples including the analysis and classification writing style of the Forsyte saga and the Lord of the Rings novels by using whole content with dynamic similarity method [16]. At the same time, current research on text classification-related tasks, such as sentiment analysis, has focused on deep learning techniques, such as Bi-LSTM, CNN, and CNN-BiLSTM (Hybrid approach) [17-19].

2.2.1. *CNN text classification.* Nowadays, CNNs are increasingly being used in text classification by converting text into numbers and manipulating them into reduced-dimensional vectors. This process, which is called "word embedding", can be used to perform analysis with the CNN technique. For example, in classifying research on Thai poems, Promrit and Waijanya [20] created a CNN deep learning model to analyze the types and attitude communication styles of these poems. They then brought the dataset to conduct word embedding using the Word2Vec technique. The Thai data set used to perform word embedding consists of five online sources for Thai content and other information related to Thai poems. The total number of words is 5.9 million words, which are then used to create a bag of words for 101,432 words. However, their CNN Thai poem sentiment analysis task only reached a low 61% accuracy. To achieve good performance for Thai text classification tasks, such as sentiment analysis, a hybrid approach, such as the Bi-LSTM CNN model with various types of embedding features, is implemented [21] to classify the sentiment of Thai content. The authors used Bi-LSTM CNN as a hybrid model with Thai-SenticNet5 corpus. For three datasets, namely, ThaiTales, ThaiEconTwitter, and Wisesight, F1-scores of 74.36%, 77.07%, and 55.21%, respectively, were achieved. However, despite the use of the Thai specific corpus and hybrid technique, the performance still did not reach 80% or above, especially for the Wisesight dataset (i.e., the Thai social media dataset). Apart from Thai content [22], the CNN text classification approach has also been incorporated for analyzing the Chinese Sogou news portal dataset by implementing word vectors based on Skip-gram models, which is a bit better than the one-hot based approach. It reached an F1-score of 97% for 10,000 news; however, when using 1,000 news for the dataset, the performance dropped significantly to 77%. From this work, we also noted that the performance of a small dataset could be increased slightly to 81% by adding another CNN layer to the two-layer CNN approach. Furthermore, using CNN with fastText as a word embedding technique for classification social networking comments has produced a more efficient and accurate CNN with Word2Vec model [23].

2.2.2. *Transformer text classification.* In [24], the principle of similarity was applied to identifying the author of the content, although Machine Learning (ML) techniques were used to build a Siamese network. The authors also compared each article in pairs with the CNN model and used Bidirectional Encoder Representations from Transformer (i.e., BERT), which is a technique developed by Google specifically for text processing. Transformer models like BERT have garnered increasing popularity in the field of text classification research due to their high performance and ability to overcome CNN weakness, which can be seen in the combination of CNN and the Transformer model called Attentive Convolutional Transformer (ACT) model [25]. The reason for building ACT is to balance the weaknesses and strengths of the CNN and Transformer model. CNN has excellent local feature extraction capability, but it has weak ability in capturing sequential and long-distance dependencies. While Transformer can solve these CNN issues, it often requires more resources, including CPU/GPU memory and computational power, especially for long-text classification that can serve as a guideline for our future research. For evaluating the ACT model, they compared ACT with three groups of text classification models: CNN and its variants, LSTM and its variants as base-line recurrent neural network, and Transformer and its variants. Results showed that ACT obtained the highest performance, while Transformer models obtained more than the accuracy of 90% for the 2-class Yelp reviews, AG's News, and the DBPedia dataset, as well as 68% for the 5-class Yelp review dataset. Furthermore, all these techniques are better than the CNN and LSTM models.

In this work, a similar situation happened in other reviewed studies that obtained a lower accuracy when using smaller datasets. However, their performance is still better than that of the CNN and LSTM models upon comparison. For Thai-language text classification, only a few studies have used a transformer-based model, and one of the possible reasons is that there are limited pre-trained transformer-based Thai-language models with relatively smaller datasets. Thus, in 2021, Thai researchers from the Vidyasirimedhi Institution of Science and Technology (VISTEC) [26] developed a large-scale Thai pre-trained transformer called WangchanBERTa based on RoBERTa, another well-known pre-trained transformer model. The VISTEC research team implemented text classification on Wisesight Sentiment dataset and obtained a micro-averaged F1-score of 76.19%, which is better than other Transformer models, such as mBERT and XLMR.

From the above literature review, we understand the drawbacks and strengths of CNN and Transformer models and use these as guidelines in implementing the experiment for our Thai dataset based on a hypothesis. Results showed that the Transformer model for Thai language with large pre-trained corpus can perform the best accuracy.

## 3. Proposed Methodology.

### 3.1. Process overview.

The process overview of the proposed methodology is displayed in Figure 1. We used the Thai-language dataset with 600 posts from the Facebook page Aizhongchina, in which each post consisted of at least 250 words. We applied the cleaning process for data preparation and then used the cleaned text as inputs to classification models that are separated into two main group: 1) the CNN Conv1D and its variant models, including the extended Conv1D by using Skip-gram and fastText Skip-gram, and 2) the Transformer and its variant models with three sub-models, namely, BERT, RoBERTa, and WangchanBERTa. There are some research works that compared CNN and Transformer models with their developed models, such as Word2Vec, fastText, and BERT, which they used with the Transformer model for comparing with the Label-Oriented (LO) approach model for text classification [27].

Our research purpose is not to find the best word embedding method but to show which classification method is the best for our timely/timeless classification based on the dataset. Thus, the chosen classification models in this research are different word embedding corpus and methods.

### 3.2. Dataset.

The Thai-language dataset was collected from the Aizhongchina Facebook page from January 2019 to October 2021. As a case study, this Facebook page posts mainly about China, has 280,000 followers, and an average total post engagement and reach of one million views per month. With permission from the page owner, the researcher retrieved data by using the manual method. The retrieved dataset contains 600 articles (at least 250 words per article) and is classified into two classes: timely and timeless.

The classification into timely and timeless categories can be described in the following steps.

1) Consider the type of article. If it is a news-type or story update about what is trending during the last seven days (e.g., COVID-19) and is widely known at a specific time, then it is classified as timely.
2) From 1), if it is a general storytelling-type of article that educates without keywords related to the current, it is classified as timeless.
3) Check whether a word is trendy in the following two ways.

First, check whether it is trending in Thailand by considering traditional media, such as TV and newspapers, and online media, such as Twitter, Facebook, Tiktok, and Google trends.

Second, check whether it is trending in China if the dataset is related to Chinese affairs and the primary audiences are interested in China. Thus, the trendy words are analyzed from both traditional media, such as TV and Chinese newspapers, and online media, such as Weibo, WeChat, Douyin, Xiaohongshu, and Baidu.

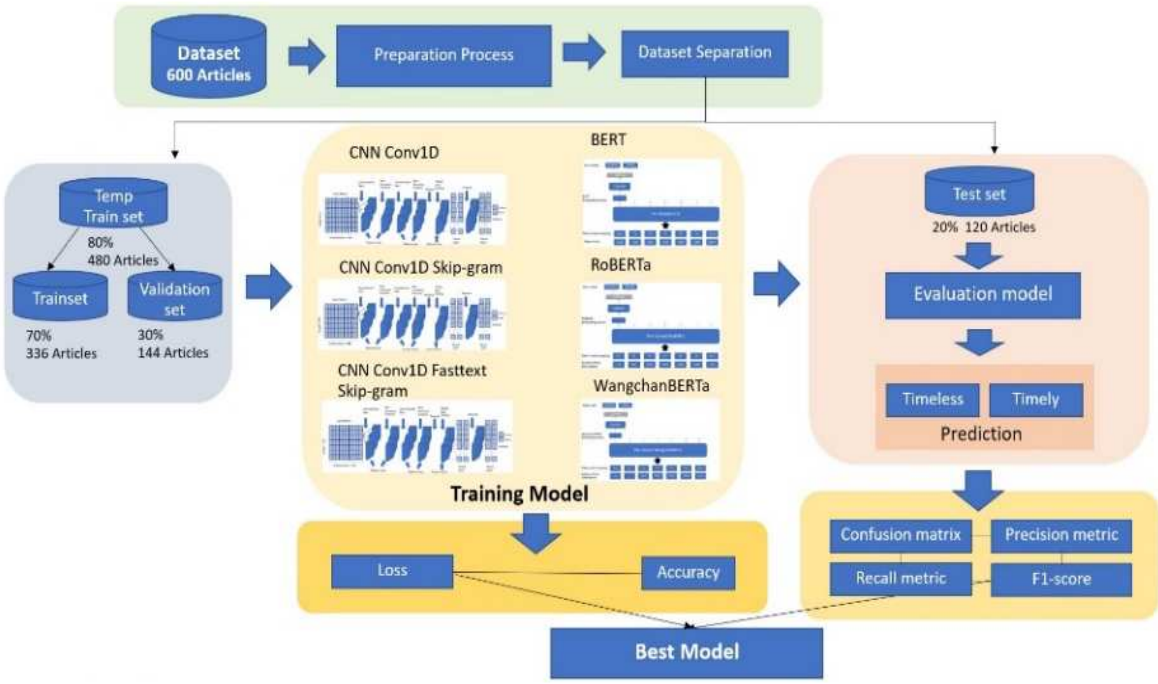



FIGURE 1. Process overview of the proposed methodology

3.3. **Data preparation.** All 600 articles were separated randomly into three sets of data: training, validation, and testing. At first, 600 articles were separated into training and testing sets with a ratio of 80 : 20 (480 articles for training and 120 for testing). Then, we separate the 480-article training set again with a ratio of 70 : 30, that is, 336 for the training set and 144 for the validation set. Each separated dataset contains two classes equally.

For preprocessing, we wrote our own functions for cleaning the data. The working principles include 1) removing any special characters and 2) removing English and Chinese characters, leaving only Thai language. This is because these are content articles related to China but written in Thai language. The primary audience is Thai, so the timeless and timely divisions are related only to Thai language. 3) Subtracting the numbers. Some studies involve other factors, such as numbers indicating time in the future. However, the time factor has not yet been included in this research; therefore, we remove the numbers and leave only Thai characters. 4) We remove the website links and hashtags because our purpose is to consider only an article's content. In CNN Conv1D and its variants models, one-hot encoding is applied into the dataset before incorporating it into the text classification model. One-hot encoding is encoding categorical features as a one-hot numeric array (0 and 1), in which the encoder derives categories based on the unique value in each feature by default [28].

3.4. **Model development.**

3.4.1. *CNN Conv1D and its variant classification models.* A CNN is a neural network inspired by the human brain and is used to visualize large areas into smaller areas for feature analysis. Before taking them together to analyze all significant areas, an essential part of CNN is feature learning, which involves the extraction of the attributes of the data toward a greater understanding of such data [29].

Figure 2 shows CNN Conv1D architecture and neural network configuration detail, which we applied in this research. We inputted the cleaned text from the data preparation process into the CNN layers, in which the first and second layers are 128 filters with a kernel size of 5 for each layer. We also set the dropout to 0.5 after the second layer and put the Global Max Pooling to reduce the overfitting issue. Then, we included a dense layer with batch normalization before passing to dropout to reduce overfitting again and going through the dense layer. Finally, we output the layer, which is determined as Softmax to classify them into timeless and timely classes. We also used the Adam optimizer.

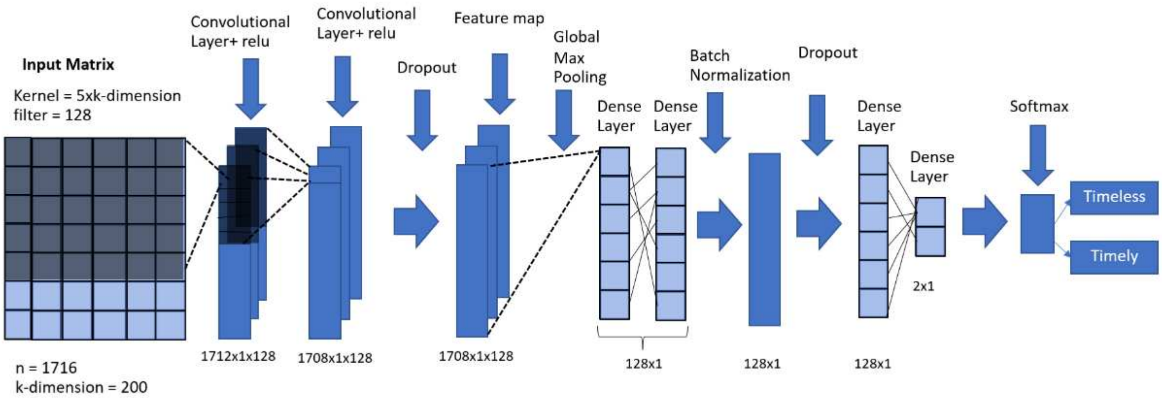



FIGURE 2. Convolutional neural network Conv1D architecture

There are two other CNN Conv1D variant classification models as follows:

- 1D CNN (Conv1D) with Skip-gram Classification Model. This model architecture is similar to CNN Conv1D but it changes the embedding matrix by using Skip-gram (Word2Vec) with dataset to make a corpus that can obtain 8,123 vocabs after processing. Skip-gram (Word2Vec) method is commonly used in text classification trends and is applied to Thai text classification (e.g., Thai poems) [20].
- 1D CNN (CNN Conv1D) with fastText Skip-gram Classification Model. This model is the same as the Skip-gram as mentioned above but it changes the corpus by using pre-trained fastText word embedding [23] as embedding matrix.

3.4.2. *Transformer classification models.* The TNN is a state-of-the-art architecture that focuses on working on Sequence to Sequence (Seq2Seq) jobs based on the core mechanism (e.g., attention mechanism). Several studies have indicated that the attention mechanism can solve bottleneck issues, such as missing data in a Seq2Seq process like translation tasks. Attention mechanism can create an output that focuses on any input position. Since the introduction of the GPT, the first pre-trained Transformer model, various pre-trained Transformer models have been trained on a large dataset based on self-supervised learning. Three transformer classification models are used in this research as follows.

- BERT. This is a trendy text processing technique with ML that Google made explicitly for text processing and can be used to process Thai language [30]. For text classification with BERT, there are special tokens [CLS] and [SEP] in the tokenization process [31] that are passed on to pre-trained BERT. The output from the pre-trained BERT is called the BERT Embedding Vector.
- RoBERTa. RoBERTa [32] is different from BERT both in terms of tokenization and token-index mapping process. It uses SentencePiece Tokenization as a mechanism, which is quite advantageous as a requirement for pre-tokinzation. Thus, the tokenization is better than BERT.
- WangchanBERTa Pre-trained Thai Language Model. WangchanBERTa [26] is explicitly developed for processing Thai datasets with large pre-trained corpus and solves the problem from RoBERTa tokenization by counting spacing, which is similar to the alphabet. Thus, there is an issue when processing Thai sentences because of the use of spacing to tokenize words. WangchanBERTa is developed based on RoBERTa so the token form is similar, although there is a specific token for spacing in WangchanBERTa as <_>. For example, inputting a Thai sentence into WangchanBERTa SentencePiece Tokenization with its rule as <s>จีนจำกัดการ เดินทางในมณฑลเหอเป่ยทางเหนือของจีน หลังพบผู้ติดเชื้อโควิด</s>, we can obtain '<s>','_','จีน','จำกัด','การเดินทาง','ใน','มณฑล','เหอ','เป่','ย','_','ทางเหนือ','ของจีน','_หลัง','พบผู้','ติดเชื้อ','โค','วิด','</s>' (The meaning of the Thai sentence is "China restricts travel in Hebei Province, North China after the discovery of local infections"). Figure 3 shows the WangchanBERTa structure, which is similar to those of BERT and RoBERTa. However, there are two differences among them: tokenization and token-index mapping, which is already mentioned in each model. All three Transformer variant models use Softmax as the classification function similar to the CNN models.

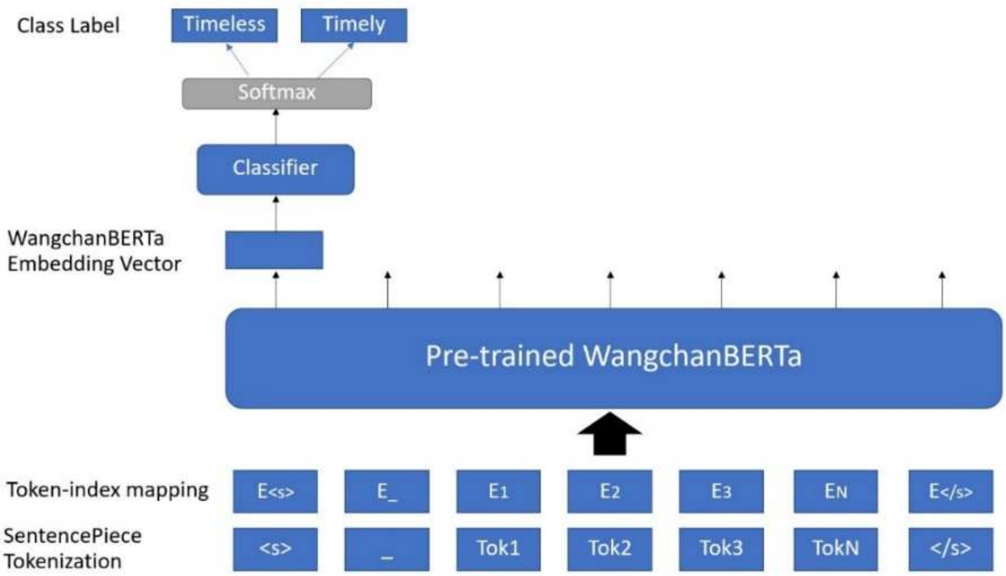



FIGURE 3. WangchanBERTa text classification structure

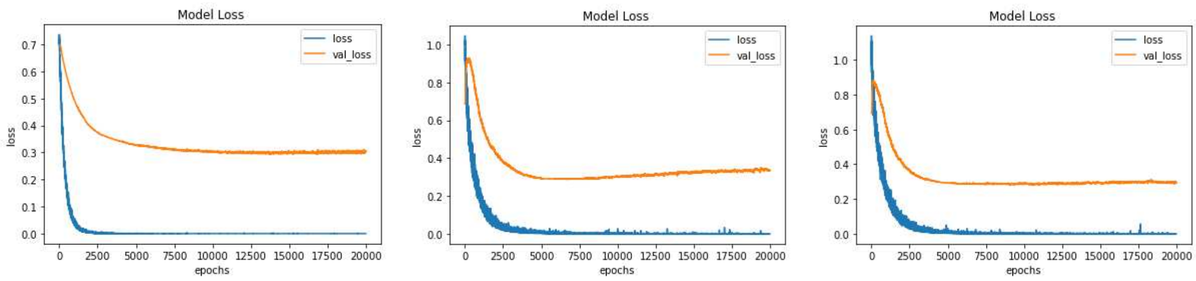
For all three Transformer models, we use the Hugging Face Transformers package to access these pre-trained models based on Transformer architecture. The reason for using Hugging Face is that this package offers pre-trained models for various tasks, including text classification (e.g., BERT, RoBERTa, and WangchanBERTa). This also offers many functions in its Transformer library to perform tasks, namely, Automodel for specific pre-trained model from Hugging Face model, and Autotokenizer for specific tokenizers from the Hugging Face pre-trained model.

3.4.3. *Hyperparameter tuning.* For setting the number of epochs, the training was done for 10,000 to 20,000 epochs for the CNN models and 10 epochs for the Transformer models. To avoid overfitting due to overtraining, we saved the checkpoints at the best epoch for the minimum validation loss. Then, when testing the trained model with the test data, we loaded the best performing one for each model. Then, we set learning rate to 0.00005 for both CNN and Transformer models due to the use of function lr_find from the Ktrain library. We also set 32 and 6 as the batch sizes for CNN and Transformers with their variant models, respectively. Other hyper-parameters for CNN and its variants, such as number of CNN layers, optimization function, and configuration of the CNN neural network, are already explained in Section 3.4.1. Moreover, for Transformer and its variant models, only the number of epochs, learning rate and batch size are set by hand.

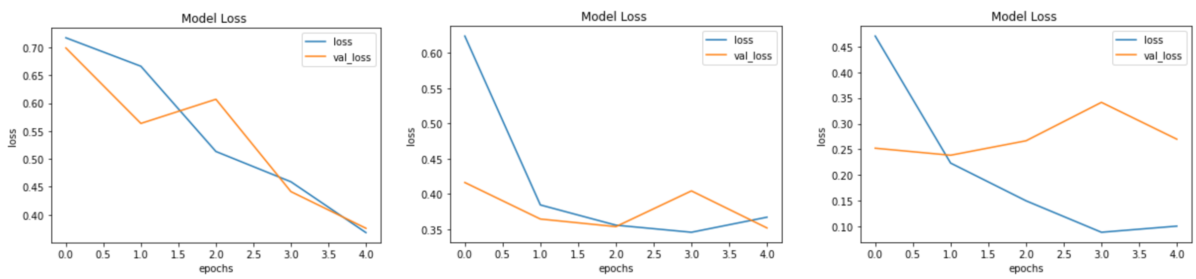
## 4. **Experiment Results and Discussion.**

4.1. **Evaluation by loss value.** The validation loss value of CNN Conv1D model as Figure 4(a) presents a slightly stable status since 10,000 epochs. There is a fluctuation after 15,000 epochs. However, validation loss is still closer to 0.3. The CNN Conv1D with Skip-gram model in Figure 4(b) is quite different from CNN Conv1D. The validation loss value presents a significant increasing trend since 6,000 epochs, and overall validation loss is lower than 0.4. The training loss of this model fluctuates as it approaches zero. The difference between training loss and validation loss of Conv1D with Skip-gram is larger than the CNN Conv1D model, and the validation loss is also slightly higher.

The validation loss of Conv1D with fastText Skip-gram, as shown in Figure 4(c) is more stable than CNN Skip-gram. It is slightly increasing since 6,000 epochs, and the overall value is below 0.4 while training loss is similar to CNN Skip-gram; there is also a slight fluctuation and the value approaches zero. The other three Transformer models: BERT, RoBERTa, and WangchanBERTa, also present a decreasing trend for loss, especially validation loss, but the training losses of RoBERTa and WangchanBERTa present an increasing trend after passing three epochs while the validation loss of their models has a decreasing trend at three epochs. However, both training and validation losses for both two mentioned models are better than those of BERT.




(a) CNN Conv1D (b) CNN Conv1D with Skip-gram (c) CNN fastText Skip-gram

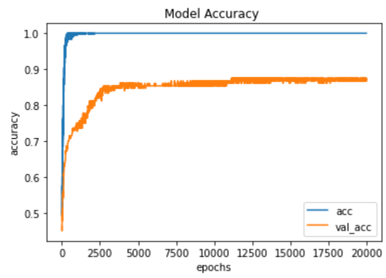



(d) BERT (e) RoBERTa (f) WangchanBERTa

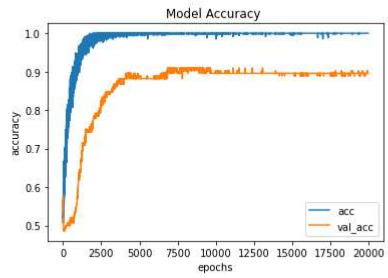
FIGURE 4. Loss of CNN Conv1D models and Transformer models

When comparing all models, the best model based on loss value is WangchanBERTa. It has the least values both in terms of validation loss and training loss, although its training loss is increasing and may lead to an overfitting issue. We present all model validation loss values in Table 1 in Section 4.3 to compare them in terms of accuracy value and confusion matrix.

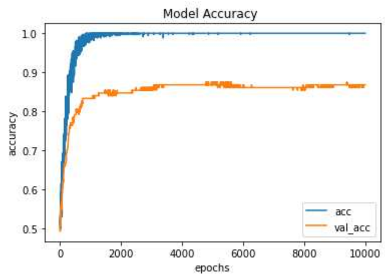
4.2. **Evaluation by accuracy value.** The accuracy value of CNN Conv1D model in Figure 5(a) has a stable trend, but there are many fluctuations given that 2,500 epochs with overall validation accuracy value is closer to 0.9. As for the accuracy value of CNN Conv1D with Skip-gram model, the training accuracy is closer to 1, which is the same as that of the CNN Conv1 model but with more fluctuations. The validation accuracy of the CNN Skip-gram, shown in Figure 5(b), is also closer to 0.9 with significant fluctuation similar to CNN Conv1D. In comparison, the validation accuracy value of the CNN Conv1D fastText Skip-gram in Figure 5(c) is around 0.9 too, with the least fluctuations compared to the CNN Conv1D and CNN Skip-gram. In Figure 5(d), the accuracy value of the BERT model with only 5 epochs is higher than 0.8 in terms of both training and validation accuracy and shows an increasing trend. RoBERTa has a decreasing trend in accuracy value both in training and validation but still closer to 0.9 and higher than BERT. WangchanBERTa, another Transformer model, also has slight fluctuation for validation at 2 and 3 epochs with the slowest point at 2. This model has the highest accuracy values of 0.95 and 0.90 for training and validation, respectively.



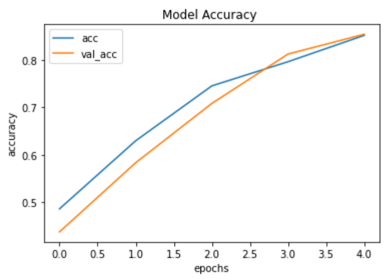

(a) CNN Conv1D



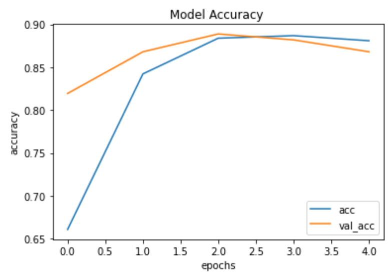

(b) CNN Conv1D with Skip-gram



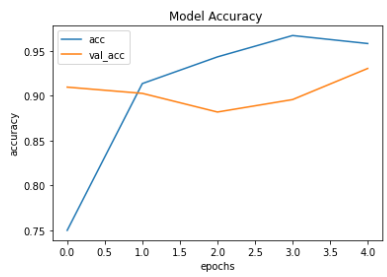

(c) CNN fastText Skip-gram




(d) BERT




(e) RoBERTa




(f) WangchanBERTa

FIGURE 5. Accuracies of the CNN Conv1D models and Transformer models

When comparing all models, the best model based on accuracy value is still Wangchan-BERTa, which has the highest validation accuracy, although its training accuracy value seems to be decreasing below 1.0 (close to 0.95). In comparison, the training accuracy of CNN Conv1D is close to 1.0. However, CNN Conv1D validation value is below 0.9, thus indicating overfitting due to the slight difference between training and validation accuracy value. We place all model validation accuracy values in Table 1 in Section 4.3 to compare the loss value and confusion matrix.

4.3. **Evaluation of text classification models.** According to the data preparation section, the data are separated into three datasets: training, validation, and testing.

The training and validation datasets are used for the training process and loss-accuracy evaluation. Thus, this evaluation is done via the confusion matrix section and uses the test dataset. Table 1 presents the text classification models' performance by calculating precision, recall, and F1-score along with validation accuracy and loss validation. As can be seen, the models with the most to least test accuracy rates in terms of text classification for two classes, timeless and timely, are WangchanBERTa, fastText Skip-gram, RoBERTa (with these models are equal), CNN Conv1D, CNN Conv1D Skip-gram, and BERT.

Table 1. Performance of the text classification models

| Model | Accuracy (Validation) | Loss (Validation) | Accuracy (Test) | Precision (Test) | Recall (Test) | F1-score (Test) |
|---|---|---|---|---|---|---|
| CNN Conv1D | 86.80% | 29.38% | 89.17% | 89.26% | 89.17% | 89.16% |
| CNN Skip-gram | 88.89% | 28.77% | 88.33% | 89.03% | 88.33% | 88.28% |
| CNN fastText Skip-gram | 88.19% | 28.06% | 90.00% | 91.14% | 90.00% | 89.93% |
| BERT | 81.00% | 40.00% | 80.00% | 80.00% | 80.00% | 80.00% |
| RoBERTa | 88.89% | 35.20% | 90.00% | 90.00% | 90.00% | 90.00% |
| WangchanBERTa | **93.06%** | **23.85%** | **93.00%** | **93.00%** | **93.00%** | **92.00%** |

5. **Conclusion.** This research aims to find the best model for text classification processing on a small Thai dataset and obtain a timely-timeless article classification model. Our results show that WangchanBERTa has the best performance for both validation accuracy and testing accuracy using only five runs. In contrast, CNN has nearly 90% validation accuracy and testing accuracy, especially when using a Skip-gram with a fastText pre-trained corpus. It has achieved a testing accuracy of more than 90%. We believe utilizing pre-trained corpus can lead to better performance.

Therefore, WangchanBERTa has the best classification efficiency than the other model tested in this research, which is the same as our hypothesis. Furthermore, the WangchanBERTa tokenizer has a vocabulary size of 25,000 sub-words, trained with 15 million Thai sentences, which is thus larger than other Transformer models focusing on the Thai language. Thus, WangchanBERTa has the best performance among all the models tested. However, we believe that creating specified pre-trained corpus related to the dataset, such as the specified pre-trained corpus for content marketing purpose including timely and timeless content, will increase performance. This will constitute our future research plan. Other factors that may possibly affect the performance of this classification, such as the number of days-time appearing in the article, which were excluded in this research, may also be included in our future research.

## REFERENCES

[1] C. Conforti, M. T. Pilehvar and N. Collier, Towards automatic fake news detection: Cross-level stance detection in news articles, *Proc. of the 1st Workshop on Fact Extraction and VERification (FEVER)*, pp.40-49, 2018.
[2] A. Hanlon, *Digital Marketing: Strategic Planning and Integration*, Sage Publications, London, 2019.
[3] L. Busche, *Powering Content: Building a Nonstop Content Marketing Machine*, O'reilly Media, CA, USA, 2017.
[4] G. Flashman, *Powerful B2B Content: Using Brand Journalism to Create Compelling and Authentic Storytelling*, Kogan Page Publishers, London, 2020.
[5] J. Smith, Digital marketing for businesses in easy steps, in *In Easy Steps*, Warwickshire, 2020.
[6] T. L. Tuten, *Social Media Marketing*, Sage Publications, London, 2020.
[7] J. R. Gallagher, *Update Culture and the Afterlife of Digital Writing*, University Press of Colorado, CO, USA, 2020.
[8] B. C. Henricks and R. Shelton, *Mastering the New Media Landscape: Embrace the Micromedia Mindset*, Berrett-Koehler Publishers, 2016.

[9] P. G. Clampitt, *Social Media Strategy: Tools for Professionals and Organizations*, SAGE Publications, London, 2017.
[10] L. Wilson, *30-Minute Website Marketing: A Step by Step Guide*, Emerald Group Publishing, United Kingdom, 2019.
[11] M. Beverland, *Brand Management: Co-Creating Meaningful Brands*, SAGE Publications, 2021.
[12] A. Gonçalves, *Social Media Analytics Strategy: Using Data to Optimize Business Performance*, Apress, Las Vegas, NV, 2017.
[13] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu and J. Gao, Deep learning-based text classification: A comprehensive review, *ACM Computing Surveys (CSUR)*, vol.54, no.3, 62, 2021.
[14] M. Javed, A. Akhtar and A. K. Yusufzai, Classification of web pages as evergreen or ephemeral based on content, *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, pp.1381-1385, 2015.
[15] S. M. Babapour and M. Roostaee, Web pages classification: An effective approach based on text mining techniques, *2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI)*, pp.0320-0323, 2017.
[16] K. Amelin, O. Granichin, N. Kizhaeva and Z. Volkovich, Patterning of writing style evolution by means of dynamic similarity, *Pattern Recognition*, vol.77, pp.45-64, 2018.
[17] R. D. Deshmukh and A. Kiwelekar, Deep learning techniques for part of speech tagging by natural language processing, *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, pp.76-81, 2020.
[18] S. Chotirat and P. Meesad, Part-of-speech tagging enhancement to natural language processing for Thai wh-question classification with deep learning, *Heliyon*, vol.7, no.10, e08216, 2021.
[19] R. Arreerard and T. Senivongse, Thai defamatory text classification on social media, *2018 IEEE International Conference on Big Data, Cloud Computing, Data Science and Engineering (BCD)*, pp.73-78, 2018.
[20] N. Promrit and S. Waijanya, Convolutional neural networks for Thai poem classification, *Advances in Neural Networks – ISNN 2017 (Lecture Notes in Computer Science)*, pp.449-456, 2017.
[21] K. Pasupa and T. S. N. Ayutthaya, Hybrid deep learning models for Thai sentiment analysis, *Cognitive Computation*, vol.14, no.1, pp.167-193, 2022.
[22] W. Xu, H. Huang, J. Zhang, H. Gu, J. Yang and G. Gui, CNN-based skip-gram method for improving classification accuracy of Chinese text, *KSII Transactions on Internet and Information Systems*, vol.13, no.12, pp.6080-6096, 2019.
[23] S. Mestry, H. Singh, R. Chauhan, V. Bisht and K. Tiwari, Automation in social networking comments with the help of robust fastText and CNN, *Proc. of the 1st International Conference on Innovations in Information and Communication Technology*, Chennai, India, 2019.
[24] C. Saedi and M. Dras, Siamese networks for large-scale author identification, *Computer Speech and Language*, vol.70, 101241, 2021.
[25] P. Li, P. Zhong, K. Mao, D. Wang, X. Yang, Y. Liu, J. Yin and S. See, ACT: An attentive convolutional transformer for efficient text classification, *Proceedings of the AAAI Conference on Artificial Intelligence*, vol.35, no.15, pp.13261-13269, 2021.
[26] L. Lowphansirikul, C. Polpanumas, N. Jantrakulchai and S. Nutanong, WangchanBERTa: Pretraining transformer-based Thai language models, *arXiv.org*, arXiv: 2101.09635, 2021.
[27] M. H. Nguyen, A label-oriented approach for text classification, *International Journal of Innovative Computing, Information and Control*, vol.16, no.5, pp.1593-1609, 2020.
[28] S. Bagui, D. Nandi and R. J. White, Machine learning and deep learning for phishing email classification using one-hot encoding, *Journal of Computer Science*, vol.17, no.7, pp.610-623, 2021.
[29] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj and D. J. Inman, 1D convolutional neural networks and applications: A survey, *Mechanical Systems and Signal Processing*, vol.151, 107398, 2021.
[30] J. Devlin, M. W. Chang, K. Lee and K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, *Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, MN, USA, pp.4171-4186, 2019.
[31] W. Hou and Y. Chen, CAUnLP at NLP4IF 2019 shared task: Context-dependent BERT for sentence-level propaganda detection, *Proc. of the 2nd Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda*, pp.83-86, 2019.
[32] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer and V. Stoyanov, RoBERTa: A robustly optimized bert pretraining approach, *arXiv.org*, arXiv: 1907.11692, 2019.