JOIV : Int. J. Inform. Visualization, 8(3) - September 2024 1213-1221

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Indonesian Fake News Classification Using Transfer Learning in CNN and LSTM

Tohpatti Crippa Praha [a], Widodo [a], Murien Nugraheni [a,*]

[a] *Universitas Negeri Jakarta, Pulo Gadung, Jakarta Timur, Indonesia*

*Corresponding author: *tohpatti@gmail.com*

***Abstract*—Fake news spreads quickly and is challenging to stop due to the ease of accessing and sharing information online. Deep learning techniques are a method that can be used to identify fake news quickly and accurately. The types of neural networks commonly utilized in deep learning architectures include Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM), which can perform well when managing the task of classifying fake news, according to several pertinent studies. Regarding handling instances of Indonesian fake news classification, this study compares how well the CNN and LSTM models perform. However, given that Indonesian is a low-resource language with scant documentation, it is challenging to build an adequate data set. At the same time, the CNN and LSTM classification models require significant training data. We proposed a transfer learning method by combining two classification models with a pre-trained IndoBERT language model. 1340 news text data were used, including 643 actual news texts from CNN Indonesia, Liputan6, and Detik and 697 fake news texts from TurnBackHoax. As a result, the performance of the combination of the LSTM classification model with IndoBERT outperformed that of the CNN classification model with IndoBERT, which only produced an accuracy of 92.91%, down by 6%, and was able to produce an accuracy of up to 97.76%, an increase of 4.8% from before. Furthermore, the results show that the LSTM classification model outperforms the CNN classification model in capturing the representation created by IndoBERT. Additionally, these insights may serve as a basis for future research on identifying fake news in Indonesia, helping to improve methods for combatting misinformation in Indonesia.**

***Keywords*—Fake news classification; Indonesian language; convolutional neural networks; long short-term memory; IndoBERT.**

Manuscript received 13 Sep. 2023; revised 7 Nov. 2023; accepted 22 Dec. 2023. Date of publication 30 Sep. 2024.

I. INTRODUCTION

With the quick progress of technology, people can conduct all their activities, especially in accessing information. Everyone undoubtedly needs information to assist them in making decisions. In accessing information, social media is one of the places that is quite attractive to the millennial generation because the use of social media is considered more accessible and more practical to fulfill all the information needed [1]. As of January 2023 alone, around 167 million people in Indonesia are active social media users, equivalent to 60.4% of the country's population [2]. However, it remains unclear to what extent the facts circulating on social media can be accurately verified [3]. The easy access and dissemination of information in the digital era often encourage irresponsible people to spread fake news.

In the last five years (August 2018 - June 2023), Indonesia has grappled with over ten thousand cases of fake news dissemination. Health-related issues emerged as the most frequently targeted category, accounting for 2,293 cases [4]. Notably, misinformation surrounding the Covid-19 pandemic has been the most prevalent. For instance, there have been misleading claims suggesting that Covid-19-related deaths were attributed to drug interactions rather than the virus itself [5]. Additionally, there's a false belief that COVID-19 vaccinations are religiously forbidden (haram) [6]. These instances of misinformation have led to widespread confusion, impeding effective pandemic management.

Followed by the next category, government-related issues accounted for 2,131 cases, making it the second most frequently encountered category of fake news. Notably, one case involved a viral video of President Jokowi giving a speech in Mandarin, which turned out to be the result of misleading editing using deepfake technology [7]. The video is labeled a hoax that could cause societal misperceptions [8]. Besides, fraud-related issues, with 1,984 cases, and political issues, with 1,392 cases, also ranked high among the most frequently encountered hoax content [4].

Fake news often incorporates conspiracy theories that lack supporting evidence and include biased or misleading statements intended to shape readers' opinions toward negative assumptions or deceive the public [9], [10]. In today's digital era, the dissemination of false information has gained widespread attention. Data from the Ministry of Communication and Informatics (KOMINFO) [11] shows that between August 2018 and March 2023, there were 11.357 instances of spreading fake news. Support from the Internet and the convenience provided by social media in uploading and sharing information has accelerated the spread and viral replication of fake news. The amount of information available on social media also causes users to feel overwhelmed by the ever-increasing surge in the amount of information. Not to mention the lack of literacy amid lots of information, making it difficult for people to verify the origin, credibility, and truth of the information they receive [12].

As the global challenge of misinformation persists, recent advancements in machine learning algorithms for detecting fake news showcase promising developments. In the context of fake news detection in Indonesia, a subsequent study advanced this endeavor, achieving an accuracy of 83.55% with the Support Vector Machine while underscoring the significance of credible sources [13]. Expanding on these initiatives, a comparative study delved into both machine learning and deep learning techniques, emphasizing the superiority of models like LSTM, which averaged an accuracy of 94.21% [14]. Contributing to the ongoing discourse, another study highlighted the effectiveness of a hoax analyzer for Indonesian news, employing the 1D-CNN model to attain an accuracy of 97.9% [15]. Seamlessly aligning with this trajectory, a subsequent investigation homed in on detecting fake news in Bahasa Indonesia, utilizing deep learning models to achieve a remarkable F1-score of 97.30% with the LSTM and Word2vec CBOW model and an accuracy of 98.38% [16]. This evolution is contrasted with a parallel study addressing COVID-19 misinformation, accentuating the efficacy of transformer-based models, particularly the CT-BERT+BiGRU model, which reached a state-of-the-art F1 score of 98% [17].

These studies have assessed the effectiveness of automatically identifying fake news using deep learning techniques. This approach can overcome the community's limitations in recognizing fake news quickly and accurately, as well as being a solution to prevent the spread of fake news. Deep learning refers to the use of neural networks with many hidden layers and is a subset of machine learning [18]. Convolutional neural networks (CNN) and recurrent neural networks (RNN) are two examples of the many deep learning architecture types that are frequently employed in NLP tasks [19]. Even though CNN is frequently employed for image-processing issues [20], in some cases, it can compete with the performance of the other models with its ability to extract spatial features from data [3]. In addition, even though the RNN can study data sequentially over a time series, it has limitations in capturing long-term dependencies, so it often experiences gradient vanishing problems [21]. Therefore, we use a modified RNN with a gate mechanism that can manage this vanishing gradient problem, called LSTM [22].

In NLP research, the availability of corpora has a major influence on producing satisfactory performance on the model. The larger the corpora, the more influential the learning model will be in recognizing language diversity. However, Indonesians are considered under-represented [23] and is known as a language with limited resources or low resource language [24]. Limited Indonesian language corpora data and resources, especially the small Indonesian language fake news dataset, can hinder research and affect model performance since large datasets are needed for deep learning models like CNN and LSTM. Meanwhile, building a new Indonesian corpus required enormous resources and effort, as well as high-performance computers. Therefore, to overcome this, a transfer learning method using a pre-trained language model is suggested [25]. Transfer learning allows existing knowledge from prior learning to be transferred to other models of completing similar tasks [26], [27]. Thus, transfer learning is an effective solution when resources are limited, as it leverages existing knowledge in pre-trained language models to allow us to prototype models quickly and reduce the required training data.

While a pre-trained language model provides a solid starting point, tailoring it to a specific language can significantly enhance contextual understanding. In this study, we employed a pre-trained language model meticulously trained on an extensive Indonesian language dataset. This deliberate choice ensured a nuanced comprehension of the Indonesian linguistic context, bolstering our approach to natural language processing tasks in the Indonesian domain. Currently, there are two monolingual IndoBERT models, namely those developed by IndoNLU [28] and IndoLEM [29], both of which are based on the BERT architecture [30]. These models have received training in the Indonesian language corpus to promote additional research on the transfer of learning to other Indonesian language processing tasks.

This study aims to evaluate the effectiveness of transfer learning in the context of Indonesian fake news classification by integrating CNN and LSTM models with the pre-trained language model IndoBERT. Our primary goal is to quantify the impact of transfer learning on the performance of the classification model, specifically assessing how leveraging IndoBERT as an embedding layer enhances the model's efficacy in discerning fake news in Indonesian texts. As IndoBERT has been pre-trained on Indonesian language resources, we aim to exploit its acquired knowledge to achieve a deeper contextual understanding of Indonesian news. This strategic utilization of transfer learning holds the potential to significantly empower the model for improved fake news detection within the Indonesian language domain.

The following are the paper's contributions:

- In this study, we introduce a novel approach by integrating the Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) architecture with the pre-trained language model IndoBERT using transfer learning.
- We investigated how the pre-trained language model IndoBERT affected the effectiveness of false news classification models in the Indonesian language.
- We presented empirical evidence comparing the performance of the proposed models in our study and the transfer learning approach with IndoBERT in classifying Indonesian fake news.

II. Materials and Method

A. Dataset

The study material comprises Indonesian-language news articles with real and fake labels. The data collection is done through the web scraping method. We use the Beautiful Soup library to pull data out of HTML and XML files. The data pulled is a collection of news website links, which, with the newspaper3k library, we extract articles and information from the news website links. As for sources of real news collection, we get them from online news portals verified by the Indonesian Press Council. The real news data was collected from CNN Indonesia, Liputan6, and *Detik* from 12 February 2023 to 14 February 2023. Meanwhile, fake news data was collected from the fact-checking website TurnBackHoax.ID in the period from August 2022 to January 2023. The data collected consists of various categories, totaling 643 real news and 706 fake news.

The selection of online news portals as sources of real news data is based on the trusted credibility of these sources so that it is considered that the news they publish is real. Meanwhile, fact-checking sites such as TurnBackHoax were chosen as the source of the fake news data because they clarified news that had the characteristics of a hoax, especially in the titles and statements they clarified. In addition, the difficulty of collecting fake news manually and the lack of adequate Indonesian language fake news data sets are other reasons for using fact-checking sites as a source of fake news data.

B. Pre-processing Data

In this study, we implemented several stages in data pre-processing, including text cleaning, by removing irrelevant special characters and symbols. Tokenization also separates text into small units, such as words. Then, each word will be checked to see whether the words are stop words, based on the list of Indonesian stop words obtained from NLTK. In addition to improving computational efficiency by reducing dataset size, removing stop words can reduce noise in Natural Language Processing (NLP) tasks by focusing on more meaningful and informative words in a text. After that, the filtered words will be stemmed to their primary form using the Sastrawi library. Stemming helps reduce redundancies in the data and provides a more consistent representation of words, thereby improving the accuracy and efficiency of the model by focusing on relevant and distinctive features.

Next, we also form a vocabulary as a list of unique words in the data, translating words into numerical representations when the model is fed with data. Then, padding and truncation strategies are used to ensure the uniform length of each text before being inserted into the deep learning model. Padding involves adding a unique token, such as a null token, to text that is too short to reach the desired length, whereas truncation involves truncating tokens to text that is too long until it reaches the specified length [31].

Following the data pre-processing step, we divide the dataset in a ratio of 60:20:20 into training data, validation data, and test data using the train-test-split method. However, when preparing data as input for a classification model, we use two different modules depending on whether the classification model uses BERT architecture. First, for the classification model without IndoBERT, we use a module from TensorFlow to tokenize data, including applying padding and truncation strategies to generate the appropriate input. Meanwhile, because models with BERT architecture have different input requirements, such as attention masks, we use a module provided by Hugging Face specifically designed to prepare input data for models with BERT architecture, such as IndoBERT.

C. Convolutional Neural Networks (CNN)

Also known as CNN or ConvNet, is a particular kind of neural network having a grid-like topological design for processing data [32]. CNN uses the mathematical operation of convolution to process data and is frequently used in the discipline of computer vision to solve problems based on images [20]. However, CNN can also be used in natural language processing to analyze text as a discrete unit [33].

In this work, 1D CNNs were employed to match the input data's dimensions [15]. In text classification, 1D CNNs can be employed to extract essential features from the text [34], like how 2D CNNs extract features from images in computer vision. In terms of computational complexity, 1D CNNs have lower complexity than 2D-CNN [35], leading to improved performance and efficiency in text classification tasks [34]Like CNN in general, 1D-CNN comprises three layers: convolution, pooling, and fully connected. Following [36], we describe how the CNN architecture works, as shown in Fig. 1.

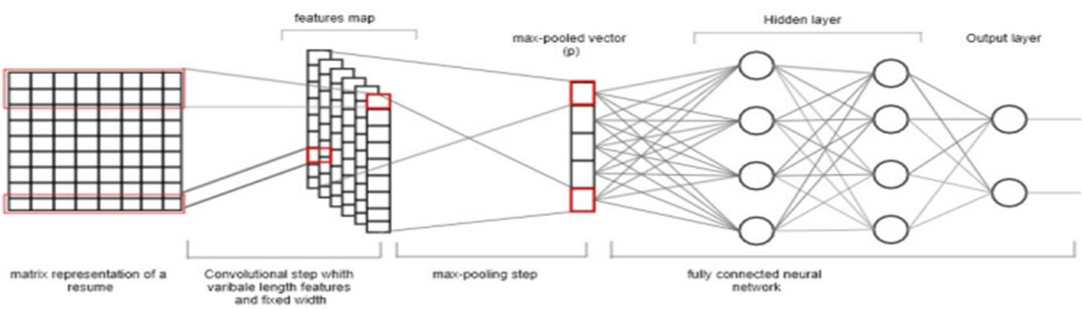

Fig. 1 1D-CNNs for text classification

The 1D-CNN architecture processes sequences of word embeddings represented as a matrix M, where each row corresponds to a word and each column represents a feature dimension. This matrix undergoes a convolution operation using learnable filters w_j of length l. This operation produces a feature map matrix F where each element $F_{(i,j)}$ represents a weighted sum of the input sequence x with the filter w_j at position i, given by the formula:

$$F_{(i,j)} = \sum_{k=0}^{l-1} x(i+k) \cdot w_j(k) \tag{1}$$

Here, the symbol $F_{(i,j)}$ denotes the value of the feature map at position (i,j). It is obtained by summing the element-wise product of the subsequence $x(i+k)$ and the corresponding weights $w_j(k)$ for k ranging from 0 to $l-1$.

This process captures interactions and non-linear patterns. After adding a bias term and applying a Rectified Linear Unit (ReLU) activation function, this step generates a feature map with dimensions $(n-l+1) \times m$, where n is the input length and m is the number of filters. Subsequently, global max-pooling is employed to reduce dimensionality, extracting the maximum value from each region to yield a d-dimensional vector.

The last step involves a fully connected layer that takes the extracted features h_i. It applies weights w_i and biases b_i to compute a weighted sum, following Eq. 1:

$$y = f(W_i \cdot h_i) + b_i, \tag{2}$$

where the activation function f is applied to determine class probabilities. In this binary classification scenario, a sigmoid activation function is utilized to classify outputs as either real or fake. The design of the CNN architecture and its shapes is further described in Figure 2.

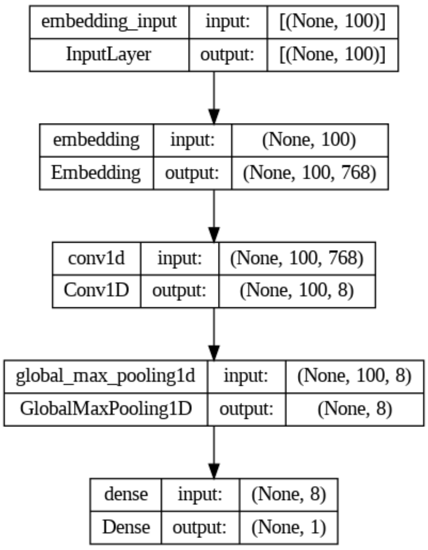

Fig. 2 The architecture of 1D-CNN classification model

D. *Long Short-Term Memory (LSTM)*

Long Short-Term Memory (LSTM) is a RNN that overcomes the vanishing gradient problem by using a gated regulator [22]. LSTM can selectively remember patterns over an extended period. Memory cells are specialized units that function similarly to accumulators or gated leaky neurons and have a connection to themselves in the following time step with a weight of one. This allows memory cells to store relevant information and accumulate external signals. However, this connection to itself is controlled by another unit which decides when the contents of the memory should be cleared [37].

Until now, LSTM has undergone many modifications and has been popularized by many researchers, resulting in various variations of LSTM. Of the many variations, the term LSTM cell usually refers to an LSTM with a forget gate [38], consisting of a forget gate, an input gate, and an output gate.

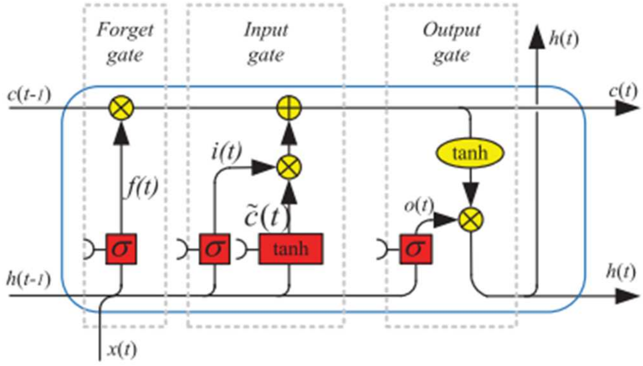

Fig. 3 The LSTM design with forget gate.

According to the connections in Fig. 3, the LSTM cell can be mathematically described as follows:

$$f_t = \sigma\left(W_{fh} h_{t-1} + W_{fx} x_t + b_f\right) \tag{3}$$

$$i_t = \sigma(W_{ih} h_{t-1} + W_{ix} x_t + b_i) \tag{4}$$

$$\bar{c}_t = tanh(W_{\bar{c}h} h_{t-1} + W_{\bar{c}x} x_t + b_{\bar{c}}) \tag{5}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \bar{c}_t \tag{6}$$

$$o_t = \sigma(W_{oh} h_{t-1} + W_{ox} x_t + b_o) \tag{7}$$

$$h_t = o_t \cdot tanh(c_t) \tag{8}$$

Then, from another source [19], it is explained that the input gate i_t, forget gate f_t, and output gate o_t determined by sigmoid (σ) functions applied over input x_t and the preceding hidden state h_{t-1}. Next, hyperbolic tangent ($tanh$) non-linearity is applied to the combined x_t and h_{t-1} to obtain a temporary result $\bar{c}_t$, which will be further utilized in the computation of the updated cell state at the current time step t. This $\bar{c}_t$ is then integrated with the historical state c_{t-1} using the input gate i_t and forget gate f_t respectively, resulting in an updated historical state c_t. Finally, the output gate o_t is applied over this updated historical state c_t to yield the final hidden state h_t.The design of the LSTM architecture and its shapes is further described in Figure 4.

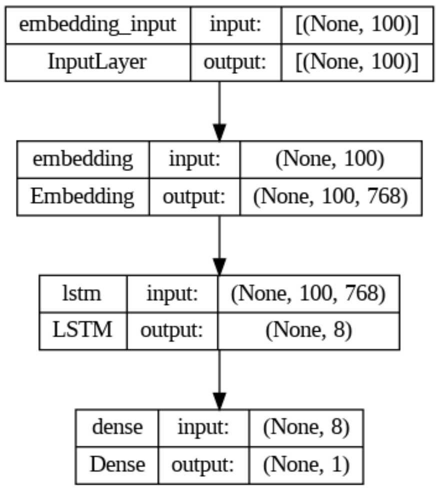

Fig. 4 The architecture of LSTM classification model

E. IndoBERT

A BERT-based language model called IndoBERT was developed utilizing corpora of the Indonesian language. To obtain the pre-trained IndoBERT model, we can train the BERT model ourselves using the Indonesian language data set, fine-tune until we get the desired performance, and then save the weights that have been learned for later use. Alternatively, we can download the available pre-trained IndoBERT models.

The IndoBERT currently comes in two varieties, suggested by the IndoNLU [28] and IndoLEM [29] teams, and trained on distinct corpora. The IndoNLU from IndoBERT is trained with around four billion words (Indo4B) from a variety of sources, including social media, web-based articles, transcriptions of video recordings, and parallel datasets [28]. Meanwhile, IndoBERT from IndoLEM was trained with more than 220 million words, which includes seventy-four million words from the Indonesian language Wikipedia, fifty-five million words from news articles, and ninety million words from the Indonesian web corpus [29].

In this experiment, we use the pre-trained language model IndoBERT, which can be acquired from https://huggingface.co/indolem/indobert-base-uncased. The IndoLEM team published this IndoBERT model [29] and is open source for anyone. The IndoBERT layer will substitute for the embedding layer to represent data in metrics.

F. Transfer Learning

Transfer learning is an approach that leverages knowledge that a pre-trained language model has learned from tasks with large data sets to use in new tasks with fewer data [26]. With transfer learning, models can start with previously learned patterns, saving time and resulting in better performance [27]. This allows using complex models even with limited amounts of data, as previously learned knowledge can be used to understand language structures and patterns in new tasks.

In our implementation, we incorporate a pre-trained model, specifically IndoBERT, as the representation layer instead of the traditional embedding layer commonly employed in NLP tasks. IndoBERT is a language model that has received training on a vast Indonesian text dataset, enabling it to capture rich linguistic patterns and contextual information specific to the Indonesian language. By leveraging IndoBERT as the new embedding layer, we hope to leverage the deep contextual knowledge and understanding of language in pre-trained models to generate highly informative and contextually rich embeddings for data input.

G. Evaluation

We use a confusion matrix to evaluate how well the built fake news classification model performed. Confusion matrices represent the model performance by generating calculations of accuracy, recall, precision, and F1 score with the equation shown as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

$$Precision = \frac{TP}{TP+FP} \quad (10)$$

$$Recall = \frac{TP}{TP+FN} \quad (11)$$

$$F1 - Score = \frac{2*Precision*Recall}{Precision+Recall} \quad (12)$$

The equation above is calculated based on four basic elements: true positive (TP), false positive (FP), false negative (FN), and true negative (TN), as described in Table 1.

TABLE I
CONFUSION MATRIX

		Predicted Class	
		True	False
Actual Class	True	True Positive (TP) The predicted fake news is true	False Negative (FN) The predicted fake news is not true
	False	False Positive (FP) The predicted real news is not true	True Negative (TN) The predicted real news is not true

News can either be detected as fake or not fake because this classification task is a binary decision problem. Additionally, we consider accuracy as a measure of model performance since precision and recall are unsuitable in this case. Meanwhile, the f1-score will later be used if there are situations where the accuracy of the model has the same results.

III. Results and Discussion

At the pre-processing stage, the data that has been collected is filtered first to ensure that the data to be processed is only news in Indonesian. The result was a decrease in the number of fake news obtained from the TurnBackHoax site from 706 to 697, while the number of real news did not change. This happened because the TurnBackHoax site received fake news reports that needed to be clarified in various languages, primarily Indonesian. So the percentage of the entire news used is 48% real and 52% fake, with 1340 news.

Next, the filtered data set is then cleaned and normalized to help remove things that are not important in the data, reducing bias that can affect model performance and increase efficiency during the learning process. We also set a maximum sentence length as input of one hundred, based on a value approximating each sentence's average maximum length on fake news labels. The reason is that the fake news collected has a shorter length than real news, so to avoid extended padding and sparse patterns, we use the average length of fake news.

After that, we divided our research into two scenarios. The first is without integrating IndoBERT into CNN and LSTM, and the second is by integrating IndoBERT using the transfer learning method in both models.

A. Classification Models without Transfer Learning

In this study, we build a simple classification model because the binary text classification task is not complex and requires a complex architecture. We define a small parameter size for the CNN classification model, with filters = 2, filters = 4, and filters = 8 in the conv1d layer. For the LSTM classification model, we also define small parameter sizes, with units = 2, units = 4, and units = 8 in the LSTM layer. Besides that, we do not apply regularization during learning to prevent overfitting. Based on the results between the CNN shown in Table 2 and the LSTM in Table 3.

TABLE II
A COMPARISON OF THE PERFORMANCE RESULTS OF THE CNN CLASSIFICATION MODEL

Model	Filters	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN	2	96.27	95.11	97.84	96.45
	4	98.51	97.2	100	98.58
	8	**99.63**	**99.29**	**100**	**99.64**

TABLE III
A COMPARISON OF THE PERFORMANCE RESULTS OF THE LSTM CLASSIFICATION MODEL

Model	Units	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
LSTM	2	83.58	81.88	87.77	84.72
	4	92.54	87.9	**99.28**	93.24
	8	**92.91**	**91.67**	94.96	**93.29**

The CNN model has better performance results than the LSTM model in classifying Indonesian fake news even though both of them were trained with a small data set so that it can be proven that the CNN model is indeed suitable for solving problems of classification tasks even in the field of NLP. Here, the CNN model can produce the best performance with 99.63% accuracy, 99.29% precision, 100% recall, and an F1 Score of 99.64%, which outperforms all the scores of the LSTM model. Meanwhile, the LSTM model was only able to produce the best performance with 92.91% accuracy, 91.67% precision, 94.46% recall, and 93.29% F1 Score. The CNN and LSTM classification results will be compared before these two classification models are applied to the transfer learning method. Furthermore, in Figure 5, we plot the CNN model's training history with accuracy metrics. Figure 6 plots the LSTM model's training history with accuracy metrics.

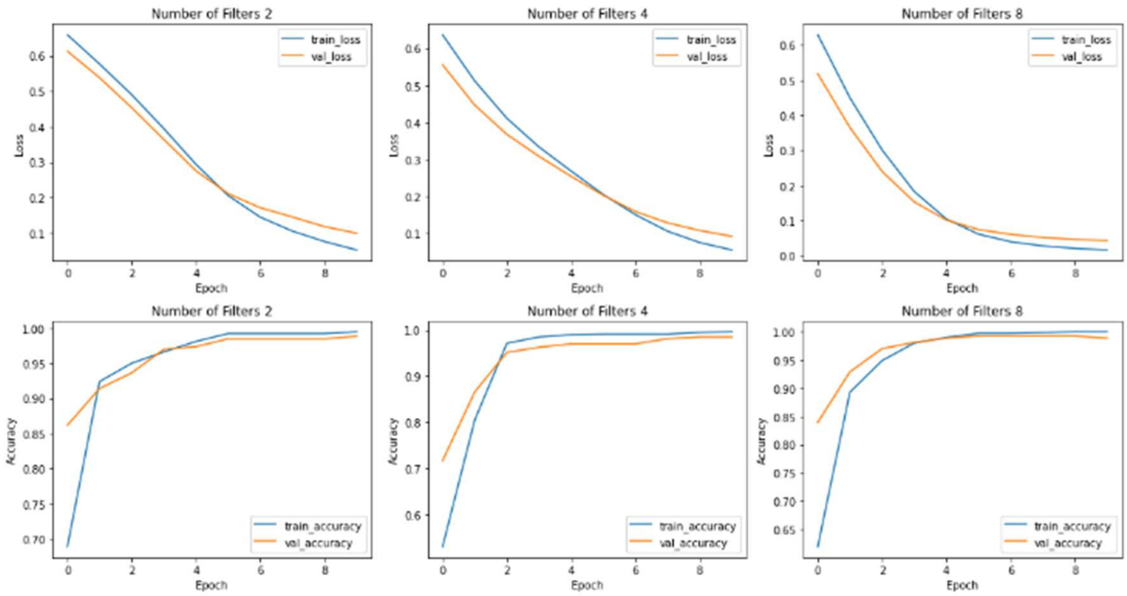

Fig. 5 CNN model training history curve with accuracy metric

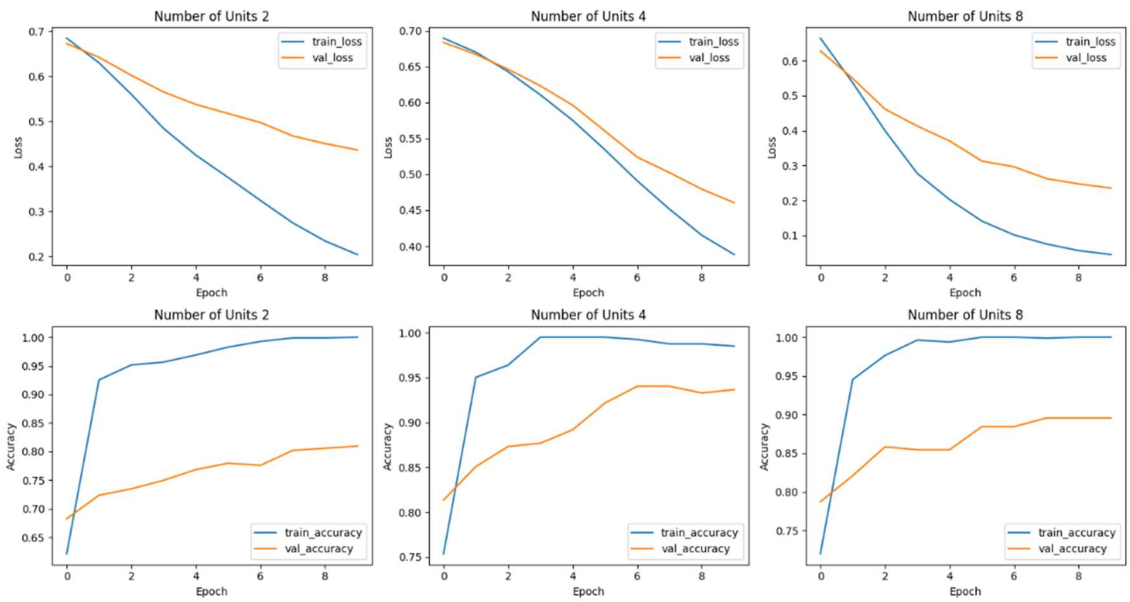

Fig. 6 LSTM model training history curve with accuracy metric

B. Classification Models with Transfer Learning

In general, text classification models use an embedding layer to represent data into metrics, enabling the model to learn unique patterns in text. However, when we apply the transfer learning method, the embedding layer will be replaced with the IndoBERT layer on top of the model, which acts as a layer to represent data into metrics. This IndoBERT layer will provide a different representation than the usual embedding layer.

In the second scenario, based on the performance result displayed in Table 4, the IndoBERT+CNN model experiences a decrease in performance compared to the CNN model without transfer learning. This could be due to the increased complexity of the model, while the model is already at the optimal point where it has produced its best performance. Increasing model complexity, but not accompanied by an increase in the quantity and quality of data, can cause the model to experience overfitting. In addition, other possibilities can also be caused by the CNN algorithm failing or not fully capturing the representation produced by the BERT layer. In more detail, we can also see the training history of the IndoBERT+CNN model in Figure 7, which shows that the model is slightly overfitting.

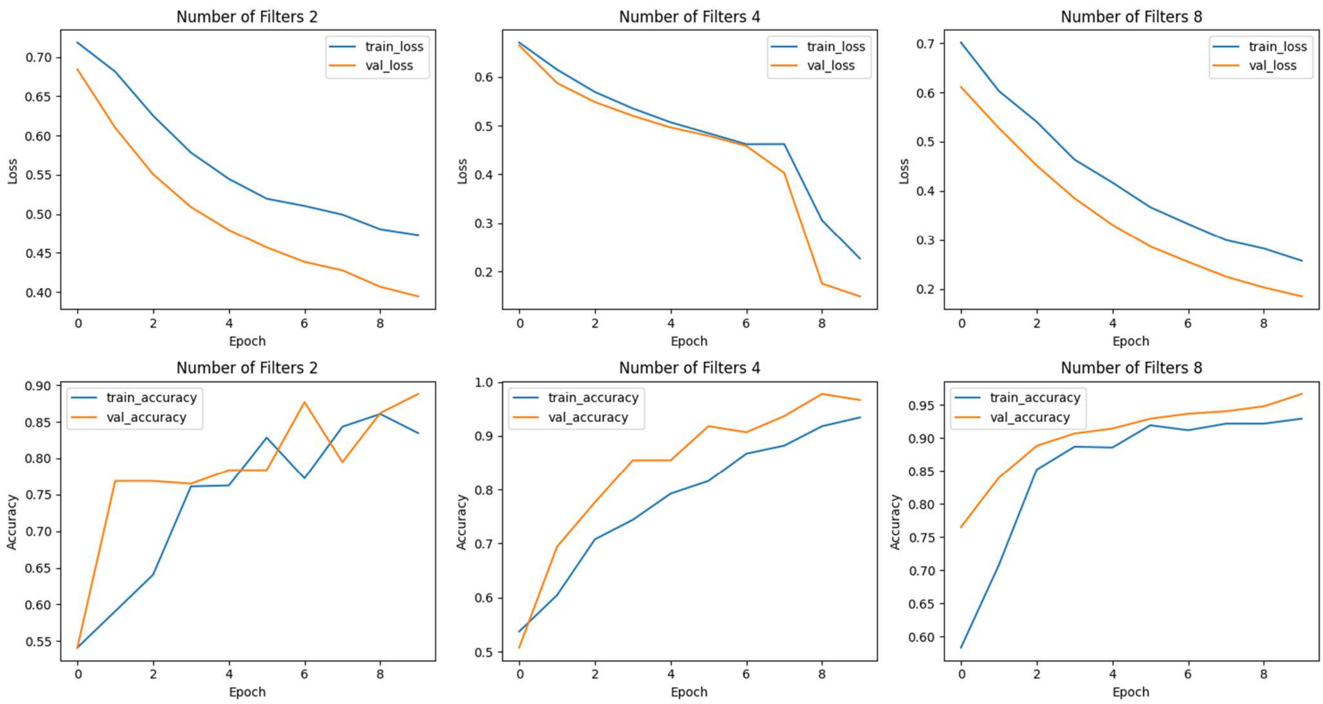

Fig. 7 IndoBERT+CNN model training history curve with accuracy metric

However, the IndoBERT+CNN combination model still produces entirely satisfactory performance. Each model with different filters has its highest score; the model with filter size 4 produces the highest accuracy and F1 score, reaching 96.64% and 96.82%, which decreased by 3% and 2.8%. Meanwhile, the highest precision was produced by the model with a filter of size 8, reaching 95.14%, which decreased by 4.15%, and the highest recall was produced by the model with a filter of size 2, gaining 98.56%, which declined by 1.44%. So, based on this analysis, the IndoBERT+CNN model with a filter size of 4 generally has the best performance among the others.

TABLE IV
A COMPARISON OF THE PERFORMANCE RESULTS OF THE INDOBERT-CNN CLASSIFICATION MODEL

Model	**Filters**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
IndoBERT-CNN	2	92.16	87.34	**99.28**	92.93
	4	**96.64**	95.14	98.56	**96.82**
	8	96.64	**96.43**	97.12	96.77

On the contrary, as shown in Table 5, the IndoBERT+LSTM model experienced an increase in performance compared to the LSTM model without transfer learning. The resulting increase in performance is not too significant, considering that the performance results of the LSTM model without IndoBERT are already relatively high. It also shows that the IndoBERT+LSTM model learns effectively during training and experiences a decrease in overfitting as shown in Figure 8. This shows the compatibility of LSTM with the representation produced by IndoBERT. In this case, the combination of IndoBERT+LSTM model architecture has a better generalization in managing the fake news classification task than the combination of IndoBERT+CNN model architecture. So, we can assume that the IndoBERT+LSTM combination model can capture valuable information patterns that the IndoBERT+CNN model may not be able to do. In this test, the performance of the IndoBERT+LSTM model exceeded that of IndoBERT+CNN.

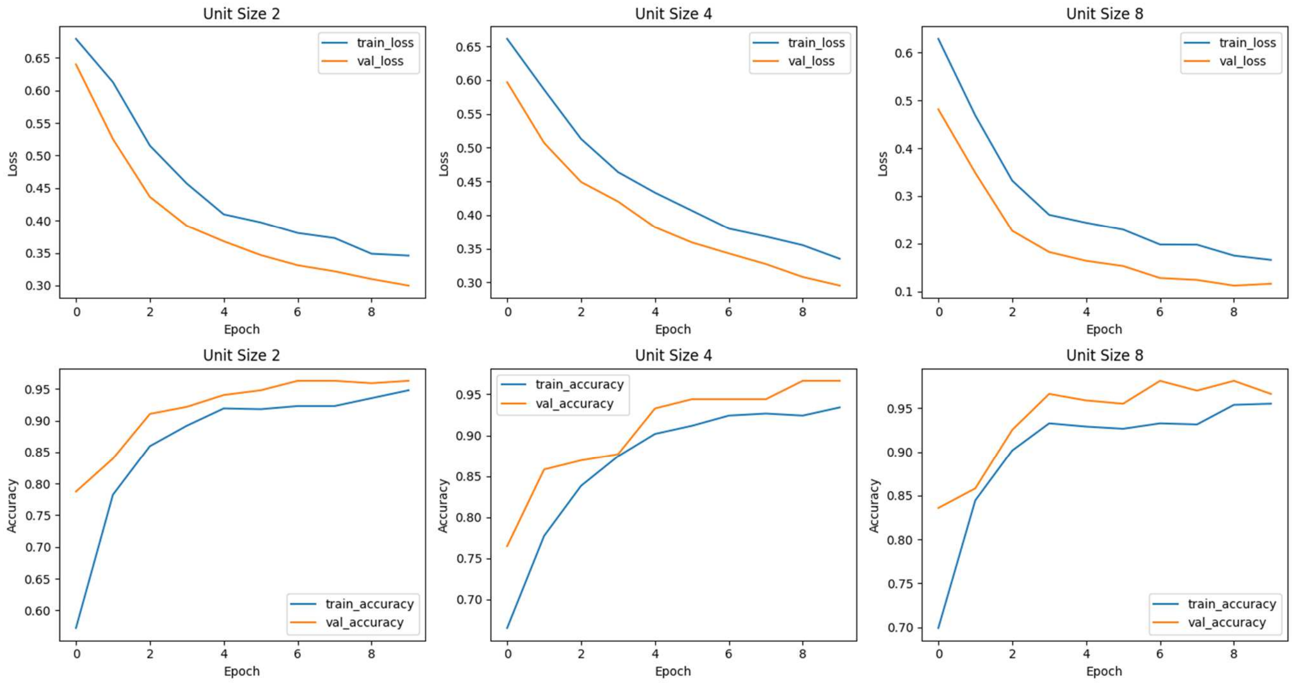

Fig. 8 IndoBERT+LSTM model training history curve with accuracy metric

Based on the analysis results, the IndoBERT+LSTM model produces an accuracy of 97.76%, which has increased by 4.85%, a precision of 96.5%, which has increased by 4.83%, a recall that has not changed and remains 99.28%, and finally, an F1 Score of 97.87%, which increased by 4.58%. Overall, this improvement occurred in the IndoBERT+LSTM model with unit size 4, and there was no apparent overfitting.

TABLE V
A COMPARISON OF THE PERFORMANCE RESULTS OF THE INDOBERT-LSTM CLASSIFICATION MODEL

Model	Units	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
IndoBERT-LSTM	2	96.64	95.77	97.84	96.8
	4	**97.76**	**96.5**	**99.28**	**97.87**
	8	96.64	94.52	99.28	96.84

C. Practical Implications of the Results

The practical implications of this research center around the potential enhancement of fact-checking effectiveness. By carefully selecting an appropriate model combination, such as LSTM and IndoBERT, we can achieve higher accuracy in verifying information. The transfer learning technique, as validated in this study, serves as a robust foundation for fortifying the model's ability to comprehend and classify information. Consequently, our research opens avenues for developing a more reliable and accurate fact-checking system, thereby contributing to the fight against the spread of fake news in the digital landscape.

IV. CONCLUSION

Based on the findings of our research, the utilization of the transfer learning method has been proven to enhance model performance. However, it's essential to acknowledge that not all model combinations are suitable, as certain algorithms may struggle to interpret the output representation produced by a pre-trained language model, potentially reducing the overall model performance. In our specific case, the fusion of the LSTM model and the IndoBERT pre-trained language model demonstrated a remarkable accuracy of up to 97.76%, reflecting a significant improvement of 4.85% from the previous performance. Conversely, combining the CNN model with IndoBERT achieved an accuracy of 96.64%, representing a decrease of 3% from the baseline. This suggests that the CNN model may not effectively capture the representation produced by the IndoBERT layer and the LSTM model.

In conclusion, developing a fake news classification model in Indonesian is an ongoing process. Future research should focus on investigating the synergistic integration of pre-trained language models with diverse neural network architectures to ensure compatibility for optimal performance. It is advisable to explore alternative methods, such as ensemble techniques, to enhance accuracy and mitigate overfitting by combining predictions from models with varying architectures. Given the pivotal role of data quantity and quality in determining model performance, researchers are encouraged to employ data enhancement techniques, particularly data augmentation, to systematically expand the training dataset and enhance the learning capabilities of the model.

REFERENCES

[1] S. C. Putri and A. Irhandayaningsih, “Literasi Informasi Generasi Millennial dalam Bermedia Sosial untuk Mengatasi Penyebaran Berita Hoax Terkait Covid-19 di Kabupaten Pati,” Anuva: Jurnal Kajian Budaya, Perpustakaan, dan Informasi, vol. 5, no. 3, pp. 491–504, Dec. 2021, doi: 10.14710/anuva.5.3.491-504.

[2] S. Widi, “Pengguna Media Sosial di Indonesia Sebanyak 167 Juta pada 2023,” DataIndonesia.id. Accessed: Jul. 09, 2023. [Online]. Available:

https://dataindonesia.id/internet/detail/pengguna-media-sosial-di-indonesia-sebanyak-167-juta-pada-2023
[3] R. K. Kaliyar, A. Goswami, and P. Narang, "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach," Multimedia Tools and Applications, vol. 80, no. 8, pp. 11765–11788, Jan. 2021, doi: 10.1007/s11042-020-10183-2.
[4] E. F. Santika, "Hoaks Soal Kesehatan Paling Banyak Ditemukan hingga Juni 2023," Databoks. Accessed: Nov. 06, 2023. [Online]. Available: https://databoks.katadata.co.id/datapublish/2023/07/21/hoaks-soal-kesehatan-paling-banyak-ditemukan-hingga-juni-2023
[5] A. Dirgantara, "Polri: dr Lois Sebarkan Berita Hoax soal COVID-19 dan Bikin Onar," detikNews. Accessed: Nov. 06, 2023. [Online]. Available: https://news.detik.com/berita/d-5640609/polri-dr-lois-sebarkan-berita-hoax-soal-covid-19-dan-bikin-onar
[6] D. Damarjati, "MUI Nyatakan Hoax terkait Virus Corona Haram," detikNews.
[7] A. S. Wardani, "Viral Video Presiden Jokowi Pidato Pakai Bahasa Mandarin, Ternyata Hasil Editan Deepfake!," Liputan6. Accessed: Nov. 06, 2023. [Online]. Available: https://www.liputan6.com/tekno/read/5433899/viral-video-presiden-jokowi-pidato-pakai-bahasa-mandarin-ternyata-hasil-editan-deepfake?page=3
[8] Liputan 6, "Komunikolog Sebut Video Hoaks Bahasa Mandarin Jokowi Timbulkan Mispresepsi," Liputan6. Accessed: Nov. 06, 2023. [Online]. Available: https://www.liputan6.com/news/read/5433695/komunikolog-sebut-video-hoaks-bahasa-mandarin-jokowi-timbulkan-mispresepsi?page=2
[9] H. Allcott and M. Gentzkow, "Social Media and Fake News in the 2016 Election," Journal of Economic Perspectives, vol. 31, no. 2, pp. 211–236, May 2017, doi: 10.1257/jep.31.2.211.
[10] A. C. Brisola and A. Doyle, "Critical Information Literacy as a Path to Resist 'Fake News': Understanding Disinformation as the Root Problem," Open Information Science, vol. 3, no. 1, pp. 274–286, Jan. 2019, doi: 10.1515/opis-2019-0019.
[11] Kominfo, "Triwulan Pertama 2023, Kominfo Identifikasi 425 Isu Hoaks," Kominfo. Accessed: Jul. 09, 2023. [Online]. Available: https://www.kominfo.go.id/content/detail/48363/siaran-pers-no-50hmkominfo042023-tentang-triwulan-pertama-2023-kominfo-identifikasi-425-isu-hoaks/0/siaran_pers
[12] J. Simarmata, M. Iqbal, M. Hasibuan, T. Limbong, and W. Albra, *Hoaks dan Media Sosial: Saring sebelum Sharing*, 1st ed. Yayasan Kita Menulis, 2019. [Online]. Available: https://www.researchgate.net/publication/336320022
[13] L. Triyono, R. Gernowo, P. Prayitno, M. Rahaman, and T. R. Yudantoro, "Fake News Detection in Indonesian Popular News Portal Using Machine Learning For Visual Impairment," JOIV : International Journal on Informatics Visualization, vol. 7, no. 3, pp. 726–732, Sep. 2023, doi: 10.30630/joiv.7.3.1243.
[14] S. A. Alameri and M. Mohd, "Comparison of Fake News Detection using Machine Learning and Deep Learning Techniques," 2021 3rd International Cyber Resilience Conference (CRC), Jan. 2021, doi:10.1109/crc50527.2021.9392458.
[15] B. P. Nayoga, R. Adipradana, R. Suryadi, and D. Suhartono, "Hoax Analyzer for Indonesian News Using Deep Learning Models," Procedia Computer Science, vol. 179, pp. 704–712, 2021, doi:10.1016/j.procs.2021.01.059.
[16] E. P. Widhi, D. H. Fudholi, and S. Hidayat, "Implementation Of Deep Learning For Fake News Classification In Bahasa Indonesia," Journal Research of Social Science, Economics, and Management, vol. 3, no. 02, pp. 370–381, Sep. 2023, doi: 10.59141/jrssem.v3i02.546.
[17] J. Alghamdi, Y. Lin, and S. Luo, "Towards COVID-19 fake news detection using transformer-based models," Knowledge-Based Systems, vol. 274, p. 110642, Aug. 2023, doi:10.1016/j.knosys.2023.110642.
[18] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," Electronic Markets, vol. 31, no. 3, pp. 685–695, Apr. 2021, doi: 10.1007/s12525-021-00475-2.
[19] W. Yin, K. Kann, M. Yu, and H. Schütze, "Comparative Study of CNN and RNN for Natural Language Processing," Feb. 2017, [Online]. Available: http://arxiv.org/abs/1702.01923.
[20] O. A. Montesinos López, A. Montesinos López, and J. Crossa, "Convolutional Neural Networks," Multivariate Statistical Machine Learning Methods for Genomic Prediction, pp. 533–577, 2022, doi:10.1007/978-3-030-89010-0_13.
[21] Muhammad Ikram Kaer Sinapoy, Yuliant Sibaroni, and Sri Suryani Prasetyowati, "Comparison of LSTM and IndoBERT Method in Identifying Hoax on Twitter," Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 7, no. 3, pp. 657–662, Jun. 2023, doi:10.29207/resti.v7i3.4830.
[22] C. Feng, M. Sun, M. Dabbaghjamanesh, Y. Liu, and J. Zhang, "Advanced machine learning applications to modern power systems," New Technologies for Power System Operation and Analysis, pp. 209–257, 2021, doi: 10.1016/b978-0-12-820168-8.00007-9.
[23] A. F. Aji *et al.*, "One Country, 700+ Languages: NLP Challenges for Underrepresented Languages and Dialects in Indonesia," Mar. 2022, [Online]. Available: http://arxiv.org/abs/2203.13357
[24] Y. Wiciaputra, J. Young, and A. Rusli, "Bilingual Text Classification in English and Indonesian via Transfer Learning using XLM-RoBERTa," International Journal of Advances in Soft Computing and its Applications, vol. 13, no. 3, pp. 73–87, Nov. 2021, doi:10.15849/ijasca.211128.06.
[25] L. H. Suadaa, I. Santoso, and A. T. B. Panjaitan, "Transfer Learning of Pre-trained Transformers for Covid-19 Hoax Detection in Indonesian Language," IJCCS (Indonesian Journal of Computing and Cybernetics Systems), vol. 15, no. 3, p. 317, Jul. 2021, doi:10.22146/ijccs.66205.
[26] B. Zoph, D. Yuret, J. May, and K. Knight, "Transfer Learning for Low-Resource Neural Machine Translation," Apr. 2016, [Online]. Available: http://arxiv.org/abs/1604.02201
[27] L. Torrey and J. Shavlik, "Transfer Learning," in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, vol. 2, IGI Global, pp. 242–264, 2009.
[28] B. Wilie *et al.*, "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," Sep. 2020, [Online]. Available: http://arxiv.org/abs/2009.05387
[29] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," Nov. 2020, [Online]. Available: http://arxiv.org/abs/2011.00677
[30] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Oct. 2018, [Online]. Available: http://arxiv.org/abs/1810.04805
[31] "Padding and truncation," Hugging Face. Accessed: Apr. 29, 2023. [Online]. Available: https://huggingface.co/docs/transformers/pad_truncation
[32] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. Accessed: Apr. 16, 2023. [Online]. Available: https://www.deeplearningbook.org/
[33] J. Patterson and A. Gibson, "Convolutional Neural Networks (CNNs)," in *Deep Learning: A Practitioner's Approach*, 1st ed., M. Loukides and T. McGovern, Eds., O'Reilly Media, 2017, pp. 238–273, 2017.
[34] H. Pasbola, "Text Classification Using Deep learning Methods," Delhi Technological University, Delhi, 2023. Accessed: Aug. 10, 2023.
[35] 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, p. 107398, Apr. 2021, doi: 10.1016/j.ymssp.2020.107398.
[36] J. K. F. Flambeau and T. Norbert, "Simplifying the explanation of deep neural networks with sufficient and necessary feature-sets: case of text classification," Oct. 2020, [Online]. Available: http://arxiv.org/abs/2010.03724
[37] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
[38] Y. Yu, X. Si, C. Hu, and J. Zhang, "A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures," Neural Computation, vol. 31, no. 7, pp. 1235–1270, Jul. 2019, doi:10.1162/neco_a_01199.