Performance Analysis of IndoBERT-Lite for Indonesian Hoax News Detection

Dzikri Muhammad Raihan*[1], Hariz Farisi[2], and Buce Trias Hanggara[3]

[1,2,3] Computer Science Faculty, Brawijaya University, Malang, Indonesia
dzikriraihan@student.ub.ac.id, hariz_farisi@ub.ac.id, buce_trias@ub.ac.id

Abstract. The growth of internet access in Indonesia, reaching 72.78% in 2024 according to the Central Bureau of Statistics (BPS), has become one of the key factors behind the increasing spread of hoax news. The Ministry of Communication and Digital (Komdigi) recorded 12,547 hoax content cases from August 2018 to December 2023. However, manual verification requires substantial time and resources. This study aims to train and evaluate the IndoBERT-lite model based on the ALBERT architecture as an automated hoax news detection solution. The research encompasses the collection of 2,000 news articles, data labeling, pre-processing, tokenization using BertTokenizer, model training, and evaluation using a confusion matrix alongside analysis of memory usage, training time, and prediction speed. The results show that IndoBERT-lite achieved an accuracy and F1-score of 98%, equivalent to IndoBERT-base. IndoBERT-lite attained a prediction time of 104.73 ms per article, outperforming IndoBERT-base which required 128.29 ms. Training for IndoBERT-lite was completed in 813 seconds using 3.83 GB of RAM and 8.19 GB of VRAM, whereas IndoBERT-base required 1,199 seconds with 6.61 GB of RAM and 8.34 GB of VRAM. During prediction, IndoBERT-lite required 1.91 GB of RAM and 0.36 GB of VRAM, compared to IndoBERT-base which needed 2.31 GB of RAM and 2.11 GB of VRAM. Based on these findings, IndoBERT-lite represents an effective and efficient solution for automated Indonesian hoax news detection.

Keywords: hoax news detection, deep learning, IndoBERT-lite, ALBERT, model evaluation

1 Introduction

Technological advancements have facilitated the rapid dissemination of information via the Internet. In 2024, 72.78% of the Indonesian population had internet access [1]. However, this rapid spread of information is accompanied by the circulation of hoax news, with the Ministry of Communication and Digital (Komdigi) recording 12,547 instances of hoax content between August 2018 and December 2023 [2]. The high exposure to such hoax news also significantly influences public opinion formation by 58.7% [3].

© The Author(s) 2026
M. Murtadha et al. (eds.), *Proceedings of the 2026 11th International Conference on Sustainable Information Engineering and Technology (SIET 2026)*, Atlantis Highlights in Computer Sciences 26,
https://doi.org/10.2991/978-94-6239-828-3_21

The public requires fast and accurate news verification, which is currently performed manually by third-party fact-checkers such as Komdigi and Mafindo, demanding substantial time and resources. Transformer-based deep learning technologies such as BERT and IndoBERT-base can address this limitation. Fardhina achieved 98% accuracy using BERT [4], while IndoBERT-base reached 99.58% accuracy after retraining on an Indonesian dataset [5].

However, IndoBERT-base relies on 124.4 million parameters, incurring heavy cost in both training and classification. The IndoBERT-lite model, based on the A Lite BERT (ALBERT) architecture with only 11.7 million parameters [6, 7], can potentially deliver faster prediction with fewer resources without compromising accuracy [7]. In this study, the authors focus on the development and performance analysis of an Indonesian hoax news detection model using IndoBERT-lite. The researchers will measure the model's accuracy, training time, prediction speed, and computational resource usage of IndoBERT-lite in detecting Indonesian hoax news. All results will be compared with the IndoBERT-base model.

2 Literature Review

2.1 Related Works

Previous studies have evaluated various models for Indonesian text classification. Rahmawati, Alamsyah, and Romadhony compared IndoBERT-base, Naïve Bayes, and SVM for hoax detection, with IndoBERT-base achieving the highest accuracy (90% vs. 80.25% and 73.42%) [8]. Azizah, Cahyono, Sihwi, and Widiarto compared IndoBERT-base, ALBERT, and RoBERTa, where ALBERT, the architecture underlying IndoBERT-lite, achieved the best accuracy (87.6% vs. 86.6% and 83.3%) [9]. Wildan and Nastiti compared IndoBERT-base and IndoBERT-lite on TikTok Tokopedia review sentiment classification, achieving 97% and 94% accuracy respectively [10]. Although not hoax-specific, this offers a useful accuracy benchmark between the two models. These studies collectively motivate training and evaluating IndoBERT-lite for Indonesian hoax detection, compared against IndoBERT-base.

2.2 Bidirectional Encoder Representation from Transformer (BERT) and A Lite BERT (ALBERT)

BERT is a deep learning model built upon the encoder of the Transformer architecture, whose core component is the self-attention mechanism. The BERT model employs the Masked Language Model (MLM) to mask words in a sentence randomly and predict them by considering the words to the left and right of the masked portion. The model also implements Next Sentence Prediction (NSP) by examining two sentences and predicting whether the second sentence logically follows the first, enabling the model to understand context [11].

ALBERT is an alternative to BERT that has been optimized to utilize computational resources efficiently by reducing the number of parameters. The ALBERT model applies two techniques to reduce the number of parameters used. First, Factorized

Embedding Parameterization, which involves using a smaller embedding dimension compared to the hidden size of the encoder, which is then projected into a larger hidden layer. Second, Cross-Layer Parameter Sharing, which involves sharing the same parameter weights across encoder layers [6].

2.3 IndoBERT-base and IndoBERT-lite

Indonesian Bidirectional Encoder Representation from Transformer-base (IndoBERT-base) is a BERT-based model trained to understand the Indonesian language (Rahmawati, Alamsyah and Romadhony, 2022). Meanwhile, IndoBERT-lite is a variant of the IndoBERT model based on ALBERT. In the IndoBERT-base model, the number of parameters used is 124.4 million, whereas the IndoBERT-lite model uses only 11.7 million [7].

2.4 Performance Evaluation Metrics

Performance is measured by confusion matrix, accuracy, precision, recall, and F1-Score [8]. Efficiency is assessed through total training time, prediction speed measured as the median of 50 prediction runs on a single article, and RAM/VRAM usage observed across three phases consisting of model loading, training, and prediction. All metrics are then compared against IndoBERT-base.

3 Methodology

3.1 Data Scraping

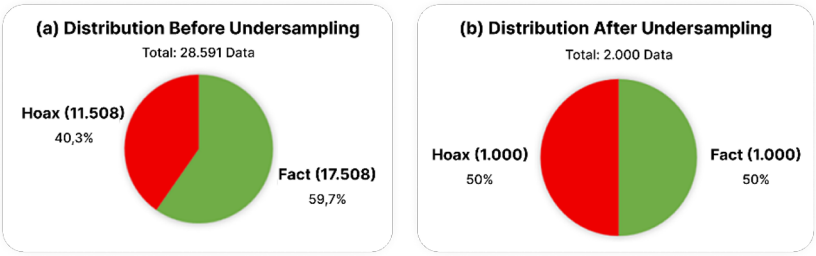


Fig. 1. Dataset Class Distribution: (a) Before Undersampling and (b) After Undersampling

The initial stage of this research involved collecting the news dataset through data scraping conducted in September 2025 from the Kompas and CNN news websites for factual news, and from TurnBackHoax and Kompas Cek Fakta for hoax news. The results, shown in Fig. 1(a), yielded a total of 28,591 news articles. However, because the obtained data was imbalanced, undersampling was performed, as shown in Fig. 1(b), by selecting 1,000 news articles from each class resulting in a dataset of 2,000 articles.

3.2 Data Labeling

Once the dataset was collected, each news article was labeled. Factual news articles from Kompas and CNN Indonesia were labeled "0", while hoax news articles from TurnBackHoax and Kompas Cek Fakta were labeled "1".

3.3 Pre-Processing Dataset

Several processes were conducted during the pre-processing stage. First, news narratives were extracted by taking the first 2-3 sentences for factual news and original narrative for hoax news. Second, data cleaning was performed to remove links, symbols, emojis, and other unnecessary characters using the Regular Expression. Third, Case Folding was applied, preventing the model from being affected by capitalization differences.

3.4 Data Split

Table 1. Data Split Distribution

Class	Training Dataset	Validation Dataset	Test Dataset
Fact (0)	700	150	150
Hoax (1)	700	150	150
Total	1,400	300	300

The dataset, totaling 2,000 articles, was divided into three parts, as shown in Table 1: 70% (1,400 articles) for training data, 15% (300 articles) for validation data, and 15% (300 articles) for testing data. This data split stage was performed using the stratified split technique to maintain class balance.

3.5 Tokenization

The tokenization stage was conducted using the BertTokenizer to convert the news text into tokens understood by the model. In addition to the tokens generated from the news input, the BertTokenizer also produces special tokens, namely the Classification Token (CLS), Separator Token (SEP), and Padding Token (PAD). Several parameter configurations in the BertTokenizer were used in this study: Max Sequence Length set to 128, Padding set to max_length, Truncation set to true, and Add Special Tokens set to true.

3.6 Model Training

Table 2. Hyperparameter Configuration Scenarios

Scenario	Batch Size	Learning Rate	Max Epochs	Optimizer
1	16	2e-5	30	Adam
2	16	3e-5		

Scenario	Batch Size	Learning Rate	Max Epochs	Optimizer
3	16	5e-5		
4	32	2e-5		
5	32	3e-5		
6	32	5e-5		

Both models were trained across several hyperparameter configuration scenarios as detailed in Table 2. A checkpoint mechanism was implemented to save the best-performing model based on validation accuracy, alongside an early stopping mechanism with a patience value of 2 to halt training if validation accuracy did not improve for 2 consecutive epochs.

3.7 Model Evaluation

Model evaluation was conducted using the 300 article testing data. The models were evaluated using several metrics: confusion matrix, accuracy, precision, recall, and F1-Score. The models were also evaluated regarding training time, prediction time, and RAM and VRAM usage when the models were loaded into the system, during training, and during prediction. All evaluation metric results were compared to determine whether the IndoBERT-lite model could achieve good and efficient performance.

4 Result

4.1 Model Training

Table 3. Training/Validation Accuracy and Loss of IndoBERT-lite

Epoch	Training Accuracy	Validation Accuracy	Training Loss	Validation Loss
1	0.9076	0.9533	0.3063	0.1413
2	0.9729	0.9633	0.0836	0.1000
3	0.9857	0.9700	0.0439	0.0818
4	0.9864	0.9367	0.0338	0.2012
5	0.9914	0.9833	0.0288	0.0598
6	**0.9964**	**0.9867**	**0.0146**	**0.0647**
7	0.9971	0.9667	0.0104	0.1194
8	0.9971	0.9733	0.0084	0.0758

After training with the configurations in Table 2, the best result was achieved with the fourth hyperparameter scenario (batch size 32, learning rate 2e-5) at epoch 6 in Table 3. Early stopping was triggered after two epochs without validation accuracy improvement, so this epoch-6 model was used as the final IndoBERT-lite model.

Table 4. Training/Validation Accuracy and Loss of IndoBERT-base

Epoch	Training Accuracy	Validation Accuracy	Training Loss	Validation Loss
1	0.9571	0.9633	0.1292	0.0747
2	0.9864	0.9567	0.0384	0.1349
3	0.9900	0.9667	0.0251	0.1092
4	0.9986	0.9733	0.0046	0.1068
5	**0.9979**	**0.9900**	**0.0033**	**0.0553**
6	1.0000	0.9900	0.0003	0.0641
7	1.0000	0.9900	0.0002	0.0656

The IndoBERT-base model achieved its best result with the second hyperparameter scenario (batch size 16, learning rate 3e-5) at epoch 5 in Table 4. Early stopping was likewise triggered after two epochs without improvement, so this epoch-5 model was used as the final IndoBERT-base model.

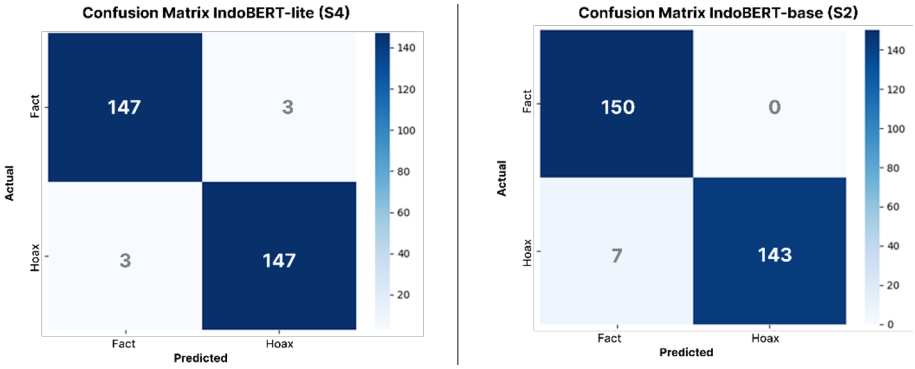


Fig. 2. Confusion Matrix of IndoBERT-lite and IndoBERT-base

Both models were evaluated using a confusion matrix in Fig. 2. IndoBERT-lite outperformed IndoBERT-base with a smaller false negative count (3 vs. 7), indicating fewer missed hoax detections.

Table 5. Accuracy, Precision, Recall, and F1-Score of IndoBERT-lite

Class	Precision	Recall	F1-Score	Accuracy
Fact (0)	0.98	0.98	0.98	0.98
Hoax (1)	0.98	0.98	0.98	

After obtaining the confusion matrix results for both the IndoBERT-lite and IndoBERT-base models, these values were used to calculate accuracy, precision, recall, and F1-Score. In Table 5, the IndoBERT-lite model achieved scores of 0.98 for precision, recall, and F1-Score for both classes. These results confirm that the model generalizes well to previously unseen data without signs of overfitting.

Table 6. Accuracy, Precision, Recall, and F1-Score of IndoBERT-base

Class	Precision	Recall	F1-Score	Accuracy
Fact (0)	0.96	1.00	0.98	0.98
Hoax (1)	1.00	0.95	0.98	

The same calculations were performed for the IndoBERT-base model, as shown in Table 6. For the fact class, the model achieved a precision of 0.96, a recall of 1.00, and an F1-Score of 0.98. Meanwhile, for the hoax class, it obtained a precision of 1.00, a recall of 0.95, and an F1-Score of 0.98. Consequently, the model achieved a final accuracy of 98%. This demonstrates that the IndoBERT-base model has an overall performance equivalent to IndoBERT-lite.

4.2 Resource Usage Analysis

Training Time

The IndoBERT-lite model was trained for 8 epochs with a total training time of 813 seconds, while IndoBERT-base required 7 epochs and 1,199 seconds. Despite needing fewer epochs, IndoBERT-base took 386 seconds longer to train than IndoBERT-lite.

Inference Speed

Table 7. Model Inference Speed

Metric	IndoBERT-lite	IndoBERT-base
Median	104.73 ms	128.29 ms
Mean	108.76 ms	126.60 ms
Std. Dev	10.01 ms	22.53 ms
Min	98.33 ms	98.70 ms
Max	138.77 ms	171.78 ms

Prediction speed was evaluated by running each model 50 times on a single test article, using the median as the primary metric for robustness to outliers. As shown in Table 7, IndoBERT-lite achieved a faster median prediction speed (104.73 ms) than IndoBERT-base (128.29 ms).

RAM and VRAM Usage

Table 8. RAM and VRAM Usage

Metric	RAM		VRAM	
	IndoBERT-lite	IndoBERT-base	IndoBERT-lite	IndoBERT-base
Load Model	1.84 GB	2.56 GB	0.35 GB	2.10 GB

Metric	RAM		VRAM	
	IndoBERT-lite	IndoBERT-base	IndoBERT-lite	IndoBERT-base
Training	3.83 GB	6.61 GB	8.19 GB	8.34 GB
Prediction	1.91 GB	2.31 GB	0.36 GB	2.11 GB

The amount of RAM and VRAM used by the models was evaluated to determine resource consumption. In this study, the available system RAM was 12.7 GB, and the VRAM from the NVIDIA T4 GPU was 15 GB. As shown in Table 8, IndoBERT-lite consistently used less RAM and VRAM than IndoBERT-base across all three phases. Most notably in VRAM during loading (0.35 GB vs. 2.10 GB) and prediction (0.36 GB vs. 2.11 GB).

4.3 Error Analysis

Table 9. IndoBERT-lite False Negative News

No	False Negative News
1	arina news. 25 jalan di jakarta bakal kena tarif. kisaran tarif sekali melintas rp5.000-rp19.900. dinas perhubungan dki jakarta bakal terapkan jalan berbayar elektronik atau electronic road pricing (erp) di sejumlah ruas jalan protokol ibu kota.
2	pesawat wing of zion, air force one milik israel yang biasa dipakai oleh perdana menteri benjamin netanyahu, lepas landas dari bandara ben gurion di tengah eskalasi konflik antara israel dan iran pada jumat (1362025). terbangnya pesawat ini memunculkan spekulasi bahwa netanyahu kabur dari negaranya atas serangan yang terjadi.
3	kabar hangat dari gedung bundar: jaksa agung st burhanuddin dikabarkan akan diganti st burhanuddin, yang menjabat sebagai jaksa agung sejak 2019, dikabarkan akan segera digantikan. menurut laporan, beliau telah berpamitan kepada internal kejaksaan agung pada sabtu, 17 mei 2025 . sosok pengganti yang disebut-sebut adalah jaksa senior dengan pengalaman di kejati dki jakarta, banten, dan sulawesi selatan.

As listed in Table 9, IndoBERT-lite yielded three false negatives on the test set. Although the two classes are generally distinguishable by style, these three hoaxes imitate factual reporting through source attribution, named entities, precise figures, and a neutral tone, so that no linguistic cue separates them from authentic news, and their falsehood resides in the claim rather than in the wording. Because each narrative falls within the sequence limit of 128 tokens, truncation is not the cause. The larger IndoBERT-base model misclassifies the same three articles among its seven false negatives, which confirms that the failures are intrinsic to the task rather than to a specific architecture, reflecting the inherent limitation of narrative-only detection and motivating fact verification through external retrieval.

4.4 Comparison with Prior Studies

Table 10. Comparison with existing Indonesian hoax news detection studies

Study	Model	Accuracy
Rahmawati et al. [8]	IndoBERT-base	90%
	Naïve Bayes	80.25%
	SVM	73.42%
Azizah et al. [9]	ALBERT	87.6%
	BERT	86.6%
	RoBERTa	83.3%
This Study	**IndoBERT-lite**	**98%**
	IndoBERT-base	**98%**

Table 10 positions our results against prior hoax and fake news studies. Both IndoBERT-lite and IndoBERT-base surpass every baseline in Rahmawati et al. [8] (IndoBERT-base, Naive Bayes, SVM) and Azizah et al. [9] (ALBERT, BERT, RoBERTa). Notably, ALBERT is the architecture behind IndoBERT-lite, yet it scored lower than our model in [9], which highlights the benefit of pretraining on the target language. As the datasets differ, the comparison is indicative rather than controlled. Nonetheless, our results are competitive, and unlike [8] and [9] we additionally provide training and inference time as well as RAM and VRAM, which constitutes a distinct contribution.

5 Conclusion and Future Work

The IndoBERT-lite model successfully detected Indonesian hoax news with an accuracy and F1-Score of 98%, a prediction speed of 104.73 ms per article, and lower resource consumption during both training and prediction, all while matching the accuracy of IndoBERT-base. For future work, researchers are encouraged to expand the dataset, explore a wider range of hyperparameter tuning scenarios to further improve model performance, and to evaluate the trained model on external news articles drawn from sources and periods outside the present corpus, in order to measure its performance on live and newly emerging hoax news.

References

1. BPS-Statistics Indonesia: Telecommunication Statistics in Indonesia (Statistik Telekomunikasi Indonesia) 2024. BPS-Statistics Indonesia. (2024).
2. Kementerian Komunikasi dan Digital: Kementerian Komunikasi dan Digital, https://www.komdigi.go.id/berita/pengumuman/detail/siaran-pers-no-02-hm-kominfo-01-2024-tentang-hingga-akhir-tahun-2023-kominfo-tangani-12-547-isu-hoaks, last accessed 2025/09/12.

3. Roy, J., Junaidi, A.: Pengaruh Terpaan Media Berita Hoax di Instagram terhadap Opini Masyarakat Milenials Akan Sumber Berita. Koneksi. 4, 280 (2020). https://doi.org/10.24912/KN.V4I2.8138.
4. Fardhina, A., Siregar, R.M., Br Sibarani, M.R.W., Br Ginting, I.C., Pratama, A.: Sistem Deteksi Berita Hoaks berbasis Algoritma Natural Language Processing (NLP) menggunakan BERT. Jurnal Manajemen Informatika, Sistem Informasi dan Teknologi Komputer (JUMISTIK). 4, 450–461 (2025). https://doi.org/10.70247/JUMISTIK.V4I1.156.
5. Millatina Aghnia Fariha, T., Lhaksmana, K.: Indonesian Fake News Classification Using Bi-LSTM With IndoBERT and Data Augmentation. ICADEIS 2025 - 2025 International Conference on Advancement in Data Science, E-learning and Information System: Integrating Data Science and Information System, Proceeding. (2025). https://doi.org/10.1109/ICADEIS65852.2025.10933014.
6. Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., Soricut, R.: ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. 8th International Conference on Learning Representations, ICLR 2020. (2020). https://doi.org/10.48550/arXiv.1909.11942.
7. Wilie, B., Vincentio, K., Winata, G.I., Cahyawijaya, S., Li, X., Lim, Z.Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., Purwarianti, A.: IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing, AACL-IJCNLP 2020. 843–857 (2020). https://doi.org/10.18653/v1/2020.aacl-main.85.
8. Rahmawati, A., Alamsyah, A., Romadhony, A.: Hoax News Detection Analysis using IndoBERT Deep Learning Methodology. 2022 10th International Conference on Information and Communication Technology, ICoICT 2022. 368–373 (2022). https://doi.org/10.1109/ICOICT55009.2022.9914902.
9. Azizah, S.F.N., Cahyono, H.D., Sihwi, S.W., Widiarto, W.: Performance Analysis of Transformer Based Models (BERT, ALBERT, and RoBERTa) in Fake News Detection. In: 2023 6th International Conference on Information and Communications Technology, ICOIACT 2023. pp. 425–430. Institute of Electrical and Electronics Engineers Inc. (2023). https://doi.org/10.1109/ICOIACT59844.2023.10455849.
10. Wildan Amru Hidayat, Nastiti, V.R.S.: Perbandingan Kinerja Pre-Trained Indobert-Base dan Indobert-Lite pada Klasifikasi Sentimen Ulasan Tiktok Tokopedia Seller Center dengan Model Indobert. JSiI (Jurnal Sistem Informasi). 11, 13–20 (2024). https://doi.org/10.30656/jsii.v11i2.9168.
11. Devlin, J., Chang, M.-W., Lee, K.,: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the 2019 Conference of the North. 4171–4186 (2019). https://doi.org/10.18653/V1/N19-1423.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

