Journal of Artificial Intelligence and Data Science
Vol. 1, No. 2, 2025, pp. 22-32
DOI:

Multimodal Hoax News Detection Using OCR and a Multi-Kernel 1D-CNN Model

Seggy Ferdianza Agustin
Program Study of Information System, Faculty of Engineering, Universitas Negeri Surabaya, Indonesia
e-mail address: seggy.22093@unesa.ac.id (corresponding author)

Received: 3 October 2025 | Revised: 12 October 2025 | Accepted: 15 November 2025
This is an open access article under the CC BY-SA license.

ABSTRACT

The focus of this research is to develop an autonomous Indonesian hoax news detection system that can leverage textual information from both news articles and text within images or videos. The proposed approach combines Optical Character Recognition (OCR) for extracting text from visual data and a Multi-Kernel One-Dimensional Convolutional Neural Network (Multi-Kernel 1D-CNN) for modeling linguistic patterns across multiple n-gram scales. The dataset is compiled from two credible Indonesian sources: Kompas.com for accurate information and TurnBackHoax.Id to hoaxes that we have verified, which sum up to 24439 labeled samples in a nearly balanced distribution of classes (50.89% valid news and 49.11% hoax news). Model performance is assessed using an 80:20 hold-out split and 10-fold stratified cross-validation. Experiments show that the proposed Multi-Kernel 1D-CNN exhibits powerful performance, but with an average accuracy of 99.918% ± 0.055, precision of 99.950% ± 0.058, recall of 99.883% ± 0.112, and F1-score of 99.917% ±0.056, which consistently outperforms the single-kernel CNN baseline method in different temporal-length test samples we have discussed above. In conclusion, the OCR and Precision–Recall curves also suggest almost perfect separability between the two classes. In summary, OCR with a Multi-Kernel 1D-CNN is an effective and efficient multimodal approach for detecting hoax news. It can be used in real-time decision-support systems for Indonesian online news.

Keywords-Hoax Detection; Fake News Classification; Multimodal OCR; Multi-Kernel 1D-CNN; Deep Learning

I. INTRODUCTION

The rapid expansion of online news portals and social media platforms has changed how people consume information, while simultaneously creating fertile ground for the spread of hoaxes, disinformation, and misinformation at scale [1, 2]. Empirical studies show that false news often spreads faster and more widely than trustworthy news, amplifying its potential to distort public opinion, undermine trust in institutions, and influence democratic processes [3]. In Indonesia, high internet penetration and intensive social media usage further exacerbate this problem, as locally tailored hoaxes exploit linguistic, cultural, and socio-political nuances that are not easily filtered by generic content-moderation tools. Manual fact-checking by journalists and volunteers, although crucial, is insufficient to keep up with the volume and velocity of misleading content, underscoring the need for automated systems to detect hoaxes.

To detect fake news mainly relied on traditional Machine Learning (ML) models, trained on hand-crafted textual features such as n-grams, sentiment scores, and stylistic indicators [4]. With the advent of deep learning, these approaches have been superseded by architectures such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Transformer-based models, which learn hierarchical representations directly from raw text [5, 6]. CNN-based models like FNDNet demonstrate that one-dimensional convolutions over word embeddings can effectively capture local n-gram patterns and yield strong performance in fake news detection [7], while hybrid models that combine BERT with CNN further enhance classification accuracy [8]. However, most of these methods remain unimodal, operating solely on textual articles or claims and ignoring the fact that a substantial portion of hoaxes today is conveyed through visual media.

Recent research on multimodal fake news detection addresses this limitation by jointly modeling text and images, often achieving higher accuracy than text-only baselines [9-11]. Frameworks such as MCred and other BERT–CNN fusion models integrate textual and visual features to capture complementary cues in social-media posts and news images [9, 10]. At the same time, Optical Character Recognition (OCR) has emerged as a critical bridge between visual and textual modalities. Many hoaxes are disseminated as screenshots, posters, or memes containing embedded text that carries the main deceptive claim. Deep Learning (DL) based OCR systems can extract this text from natural scenes, but existing multimodal fake news methods usually treat OCR output as an optional add-on or omit it entirely, focusing instead on global image features [12, 13]. Consequently, the semantic information encoded in image-embedded text is often underexploited in practical detection systems.

In the Indonesian context, several studies have explored hoax news detection using CNN and other deep learning models on Indonesian-language datasets sourced from online news portals and fact-checking sites [14-17]. Transformer-based encoders such as IndoBERT have also been shown to provide strong baselines for Indonesian fake news classification, especially when combined with downstream classifiers like LSTM or shallow neural networks [16]. Parallel to this, multi-kernel and multi-scale CNN architectures have gained attention in text classification and computer vision, as parallel convolutional filters with different kernel sizes can simultaneously capture short and long-range patterns in the input sequence [18-20]. However, to the best of our knowledge, there is still limited work that systematically integrates OCR-based text extraction from images with a Multi-Kernel CNN (MK-CNN) classifier for Indonesian hoax news, and even fewer that evaluate such architectures under rigorous cross-validation and report deployment-ready implementations.

This paper addresses these gaps by proposing a multimodal hoax news detection system for Indonesians that combines OCR with a Multi-Kernel 1D-CNN architecture. Textual inputs consist of manually provided titles and article bodies as well as OCR-extracted text from news-related images or video frames, all processed through a strict preprocessing pipeline designed to remove label-leakage cues and source-specific artifacts. The MK-CNN employs parallel convolutional filters with multiple kernel sizes to learn n-gram patterns at different granularities and is systematically compared with a single-kernel CNN baseline using both an 80:20 hold-out split and 10-fold stratified cross validation. The best-performing MK-CNN model is then deployed as a web-based application that supports real-time multimodal hoax checking and nearest-neighbour article retrieval, providing a practically proper and methodologically rigorous contribution to the detection of Indonesian hoax news.

II. METHOD

The overall research procedure follows the Cross-Industry Standard Process for Data Mining (CRISP-DM), which is widely adopted in data analytics projects to ensure a reproducible and well-documented workflow from business understanding to deployment [21]. In this study, CRISP-DM is adapted to the context of Indonesian hoax news detection. Figure 1 displays the six phases (Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment), guide the construction of the dataset, the design of the deep-learning models, and the implementation of the web-based multimodal hoax detection system at the center of the CRISP-DM cycle.

A. Research Framework

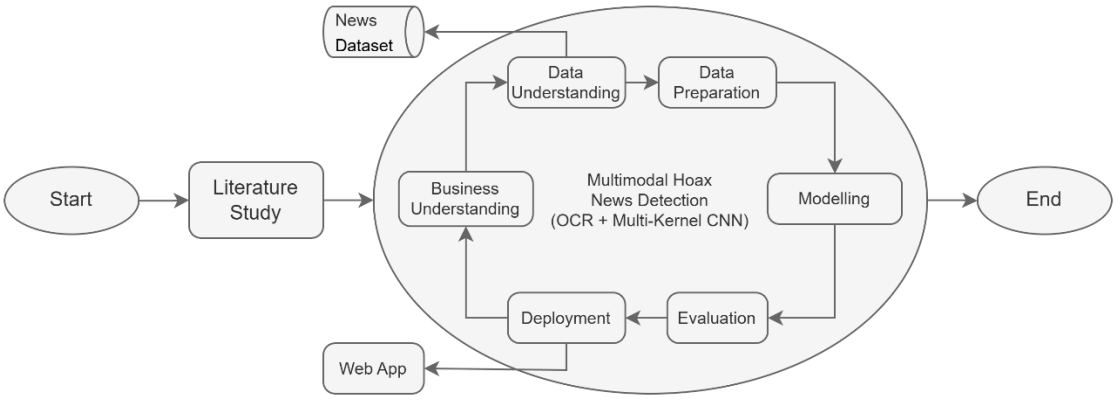


Figure. 1. CRISP-DM-based Research Framework For The Proposed Hoax Detection System.

In the business understanding phase, the main objective is to build an automated system that can detect hoax news in Indonesian by combining textual information with text extracted from images or video frames using OCR. The data understanding phase focuses on exploring the characteristics of news articles and fact-checking reports, including label distribution and potential noise. During data preparation, the raw data are cleaned, labels are standardized, and auxiliary fields

such as URLs and image links are organized. In the modeling phase, two architectures are developed: a single-kernel 1D-CNN baseline and the proposed Multi-Kernel 1D-CNN (MK-CNN), inspired by previous successes of CNN-based and multi-kernel architectures in text classification and fake-news detection [7, 18, 22]. The evaluation phase compares the two models under identical experimental conditions, while the deployment phase integrates the best-performing model into a web application that supports multimodal inputs.

B. Dataset Construction

The dataset is constructed from two complementary Indonesian sources that are commonly used in previous hoax-detection studies [23, 24]. Kompas.com provides legitimate news articles, whereas TurnBackHoax.id supplies verified hoaxes and their clarifications. Each record in the combined corpus contains: (1) a text field obtained by concatenating the news title and article body or fact-check narrative, (2) a binary label (0 for valid news, 1 for hoax news), (3) the original article URL, and (4) an optional URL to the main image or screenshot. After merging both sources, removing obvious duplicates at the URL level, and discarding incomplete entries, the final dataset consists of 24,439 labeled news items. Among them, 12,438 instances are valid news and 12,001 instances are hoaxes, resulting in a nearly balanced class distribution. The detailed statistics are presented in Table 1, which reports the count and percentage of each class as well as the overall total.

TABLE 1. DATASET DISTRIBUTION

Label	Class Diagram	Total	Percentage (%)
0	Valid News	12438	50.89 %
1	Hoax News	12001	49.11 %
Total		24439	100 %

C. Text Preprocessing and OCR Pipeline

All textual content, both native articles and OCR outputs are passed through a unified preprocessing pipeline before being fed to the models. As shown in Figure 2, raw news text or OCR text from images/videos is first concatenated into a single string. The following cleaning steps are then applied, adapted from best practices in Indonesian hoax detection and text-classification studies [23, 25]:

- Lowercasing and tag removal: text is converted to lowercase and platform-specific tags such as [HOAKS] and [SALAH] are removed.
- URL and domain filtering: explicit URLs and domain tokens (e.g., kompas.com, turnbackhoax.id) are removed to prevent the model from learning source-specific shortcuts.
- Label-cue removal: words and phrases that directly indicate the label (e.g., hoax, misinformasi, disinformasi, klarifikasi, cek fakta) are filtered out.
- Character filtering and whitespace normalization: non-alphabetic characters are stripped, multiple spaces are collapsed, and leading/trailing spaces are trimmed.
- Duplicate and empty-text handling: duplicated texts are removed and records that become empty after cleaning are discarded.

After cleaning, a word-level tokenizer is fitted on the training data only, consistent with previous CNN-based approaches to fake-news and Indonesian text classification [7, 23, 25]. The tokenizer is limited to the top 50,000 most frequent tokens; out-of-vocabulary words are mapped to a special <unk> token. Each text is then converted into an integer sequence and padded or truncated to a fixed length (e.g., 150 tokens), then producing the fixed-size input matrix for the CNN models (Figure 2).

To support multimodal inputs in the deployed system, an OCR pipeline is integrated based on deep-learning OCR techniques [14, 16]. For each image or sampled video frame, the system applies basic preprocessing (grayscale conversion, resizing, and thresholding) and then uses an OCR engine to recognize embedded text. The recognized text is processed through the same preprocessing and tokenization steps described above, then concatenated with the manually entered title and body before being fed into the classifier. When OCR output is unreliable or unavailable, the system falls back to purely textual input, ensuring robust behavior in practical usage scenarios.

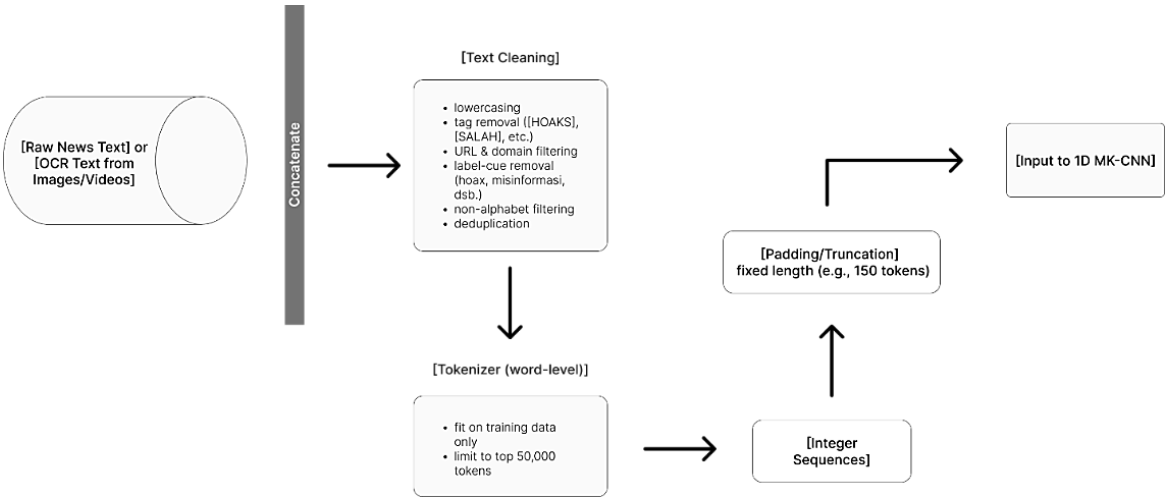


Figure. 2. Text Preprocessing and Tokenization Pipeline

D. Modeling and Evaluation Protocol

Both the single-kernel and Multi-Kernel 1D-CNN architectures share the same input representation (word embeddings over padded token sequences), but differ in their convolutional layers. The baseline employs a single 1D convolution layer with one kernel size, similar to conventional CNN models for sentence classification and fake-news detection [7, 18]. The proposed MK-CNN, motivated by multi-kernel and multi-scale CNN designs [18, 22], uses several parallel 1D convolution layers with different kernel sizes to capture short, medium, and long n-gram patterns. Feature maps from all convolutional branches are combined via global max pooling and concatenation, followed by fully connected layers and a sigmoid output neuron that produces the hoax probability. Both models are trained with the Adam optimizer and binary cross-entropy loss, using mini-batches whose sizes are adapted to available GPU resources.

The evaluation protocol consists of two stages. First, an 80:20 stratified hold-out split is used to compare the learning behavior of the single-kernel and multi-kernel models. Their training and validation loss/accuracy curves are later reported as a Graph of model training results. Second, the best-performing architecture is evaluated using 10-fold stratified cross-validation, a standard practice in text classification to obtain robust performance estimates [16]. In each fold, the tokenizer is re-fitted on the training subset, and the model is retrained from scratch. Classification metrics, accuracy, precision, recall, and F1-score, are computed on the held out fold and averaged across all folds. In addition, confusion matrices and ROC/Precision–Recall curves are generated to analyze the trade-off between true and false positives, and the Graph of training results using the proposed model is presented in the Results and Discussion section.

III. PROPOSED METHOD

This section describes the proposed multimodal hoax news detection system that combines OCR-based text extraction from images with a Multi-Kernel 1D Convolutional Neural Network (MK-CNN) classifier. Figure 3 show the architecture that start from user-provided inputs (news title, article body, and optional image or video), the system constructs a unified textual representation, processes it through a dedicated preprocessing and tokenization pipeline, and feeds the resulting sequences into the MK-CNN. The model outputs the hoax probability, which is then used both for binary classification and for providing decision support in the deployed web application.

A. System Overview

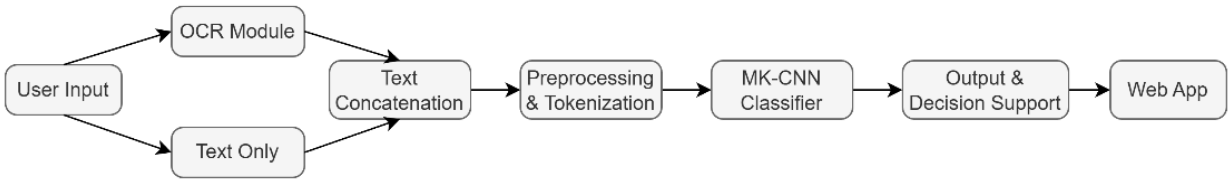


Figure. 3. Architecture Of The Proposed Multimodal Hoax News Detection System.

As shown in Figure 3, the proposed system consists of four main modules: (1) input acquisition, (2) OCR-based text enrichment, (3) text preprocessing and tokenization, and (4) MK-CNN classification with decision support. In the input

acquisition module, the user may submit a news title and body text, an image (e.g., screenshot, poster, infographic), a short video clip, or any combination of these. The OCR-based text enrichment module is responsible for extracting textual information from visual media using a deep-learning-based OCR engine [14, 16]. Selected frames from an uploaded video, or the entire uploaded image, are preprocessed (grayscale conversion, resizing, and thresholding) before being passed to OCR. The recognized texts are concatenated with the manually entered title and article body, forming a single raw text field that captures both original news content and text embedded in images or videos.

The concatenated text is then processed by the unified preprocessing and tokenization pipeline described in Section II-C (see also Figure 2). After cleaning and tokenization, the fixed-length sequence results of word indices is fed into the MK-CNN classifier. The classifier outputs a probability $p \in [0,1]$ representing the degree to which the input is predicted to be hoax. This probability is thresholded (default $\tau=0.5$) to produce the final binary label. In the deployed web interface, the prediction is complemented with a list of the most similar reference articles retrieved from Kompas.com and TurnBackHoax.id using TF–IDF and cosine similarity, which provides additional contextual information but does not affect the core MK-CNN decision.

B. Multimodal Text Construction via OCR

The key multimodal aspect of the system lies in its use of OCR to enrich textual inputs with information from images and videos. For each uploaded image, basic image processing is performed to enhance the contrast between foreground text and background, following standard practices in document and scene-text OCR [14, 15]. For videos, a small number of frames is sampled at fixed intervals; each frame is treated as an image and processed by the same OCR routine. The text recognized from all frames is concatenated, deduplicated, and passed to the text preprocessing pipeline.

By converting image-embedded text to standard textual form, the system effectively reduces the multimodal problem to an augmented text classification task while retaining the semantic content of visual hoaxes, such as captions, headlines, or overlaid slogans. This design choice allows us to reuse established NLP architectures and training procedures while leveraging the richer information in visual media. Similar ideas of exploiting OCR output have been explored in other domains, but their explicit combination with MK-CNN for hoax news detection, especially in Indonesian, remains underexplored [9, 12, 13].

C. Multimodal Text Construction via OCR

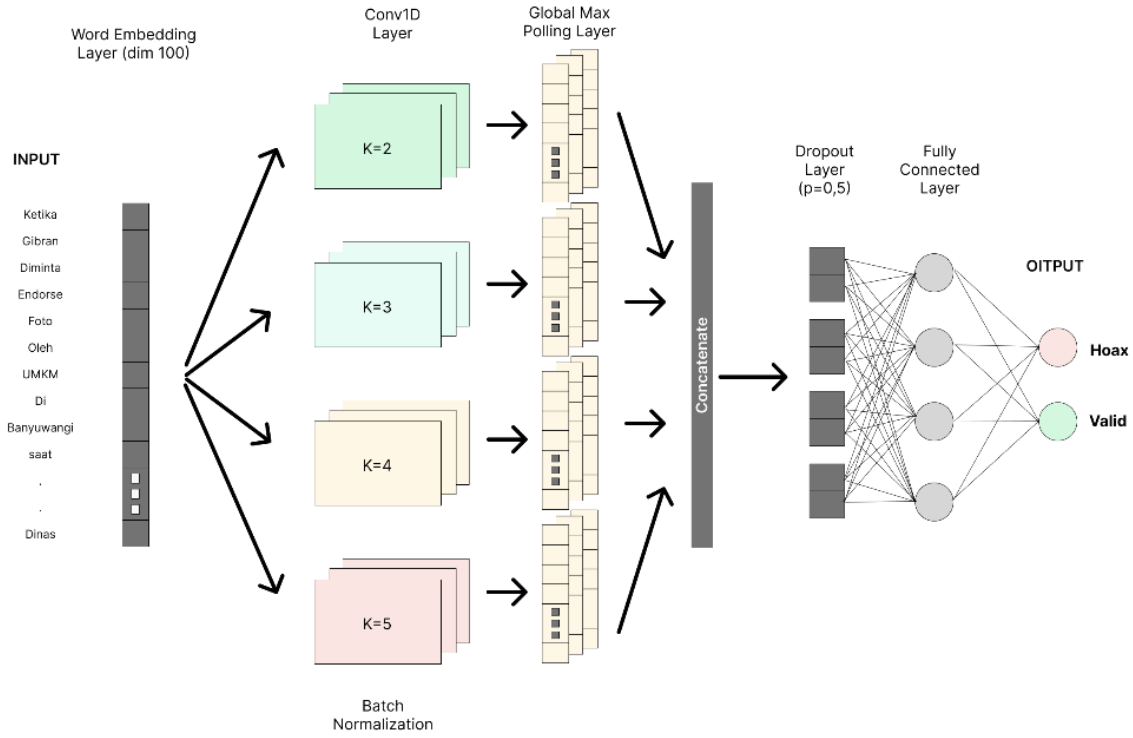


Figure. 4. Detailed architecture of the Multi-Kernel 1D-CNN classifier

Figure 4 presents the core classifier in the proposed system is a Multi-Kernel 1D-CNN (MK-CNN), which inspired by previous work on CNN-based sentence classification and multi-kernel architectures [7, 18, 22]. Let $x \in R^{L \times d}$ denote the embedded input sequence of length LLL (e.g., 150 tokens) with embedding dimension ddd. The MK-CNN applies a set of parallel 1D convolutional layers with different kernel sizes $k \in \{2,3,4,5\}$, each containing F filters (e.g., 128 filters per kernel

size). For a given kernel size k, the convolution operation produces a feature map $c^{(k)} \in R^{L-k+1c}$, which is then aggregated by global max-pooling to capture the most salient n-gram feature for that kernel size.

The pooled feature vectors from all kernel sizes are concatenated to form a single representation $h \in R^{4F}$. This vector is passed through a dropout layer for regularization, followed by one or more fully connected layers with ReLU activations. A final dense layer with a sigmoid activation function outputs the hoax probability. In contrast, the single-kernel baseline used for comparison employs only one convolutional layer with a single kernel size, thereby limiting its ability to capture multi-scale n-gram patterns [18, 20]. Previous studies have shown that multi-kernel CNNs can improve performance in various classification tasks by jointly modeling local patterns of different lengths [18-20, 22]; the present work extends this idea to Indonesian hoax news detection with multimodal text inputs.

D. Decision Threshold and Web-Based Deployment

During training, the MK-CNN is optimized using the Adam optimizer and binary cross-entropy loss. At inference time, the output probability is compared against a decision threshold τ by default τ=0.5, although alternative thresholds can be explored to balance precision and recall depending on application needs [16]. To support users in interpreting the model's prediction, the deployed web application presents not only the predicted label and probability, but also a list of top-kkk reference articles retrieved from Kompas.com and TurnBackHoax.id using TF–IDF similarit, an approach similar to previous decision-support systems for fake news verification [11, 24].

The final web-based system is implemented using a lightweight API framework that exposes the OCR and MK-CNN inference pipeline as HTTP endpoints. Uploaded text and media are processed on the server, predictions and similarity scores are computed, and results are rendered in an interactive dashboard that highlights the hoax probability, whether OCR was used, and snippets of the retrieved reference articles. This deployment demonstrates that the proposed multimodal MK-CNN model is not only effective in offline evaluations but also practical for real-time hoax screening in everyday use.

IV. RESULT AND DISCUSSION

A. Dataset Characteristics

The final dataset used in this study consists of 24439 Indonesian news items, comprising 12438 valid news articles and 12001 hoax news articles. As summarized in Table 1, the class distribution is nearly balanced, with valid and hoax labels contributing 50.89% and 49.11% of the samples, respectively. This balanced composition is advantageous because it reduces the need for additional re-sampling strategies and allows accuracy, precision, recall, and F1-score to be interpreted without strong bias toward the majority class.

(a) (b)

Figure. 5. Word Cloud for Valid (a) and Hoax (b) News

To gain a qualitative understanding of the vocabulary used in each class, separate word clouds were generated for valid and hoax news after preprocessing. Figure 5 (Word Cloud for valid vs. hoax news) show that the two classes exhibit distinct lexical patterns, reflecting differences in topics and framing commonly observed in previous fake news and misinformation studies [1, 4, 17]. These visualizations support the hypothesis that a convolution-based model operating on word sequences should be able to exploit such lexical regularities to distinguish hoaxes from legitimate news.

The first set of experiments compares the learning dynamics of the single-kernel 1D-CNN baseline with the proposed Multi-Kernel 1D-CNN (MK-CNN) on an 80:20 stratified hold-out split. Figure 6 presents the Graph of model training results for the single-kernel model, showing the evolution of loss and accuracy on both the training and validation sets across ten epochs. The curves indicate rapid convergence within the first two epochs: the binary cross-entropy loss decreases sharply and stabilizes near zero. At the same time, the accuracy quickly approaches 0.99 or higher for both the training and validation data. There is

no significant divergence between the training and validation curves, suggesting that the baseline model does not suffer from severe overfitting under the chosen hyperparameters, consistent with previous CNN-based fake news classifiers [7, 25].

The corresponding Graph of training results using the proposed model is shown in Figure 7 (loss) and (accuracy) for the MK-CNN. Similar to the baseline, the proposed model converges very quickly; however, the MK-CNN exhibits slightly lower validation loss and a marginally higher validation accuracy across epochs. The near overlap between training and validation curves again suggests good generalization on the hold-out split. These observations support the intuition from earlier multi-kernel CNN studies [18, 22] that parallel convolutional filters with different kernel sizes can extract richer n-gram features without substantially increasing the risk of overfitting, provided that appropriate regularization (dropout and early stopping) is applied.

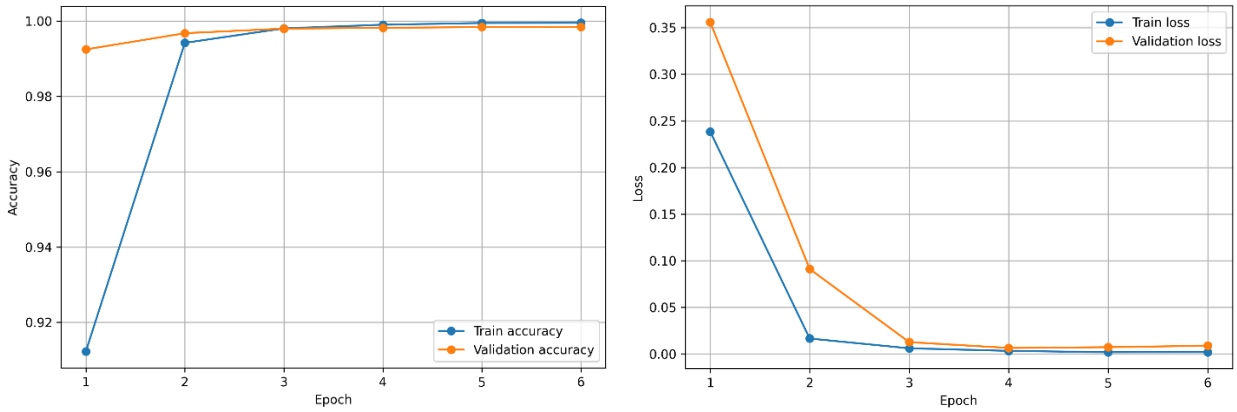


Figure. 6. Graph of Model Training Results based on Single Kernel CNN

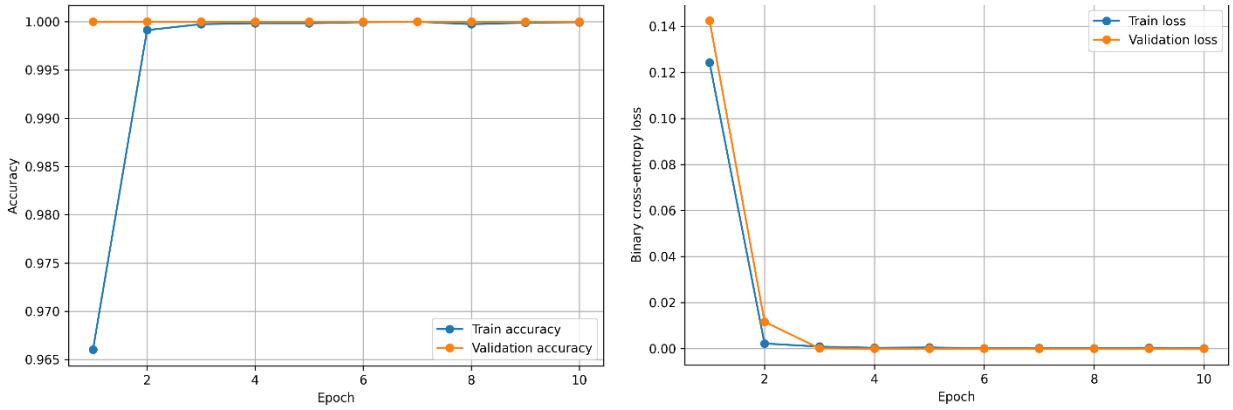


Figure. 7. Graph of Training Results Using The Proposed Model (Multi Kernel CNN)

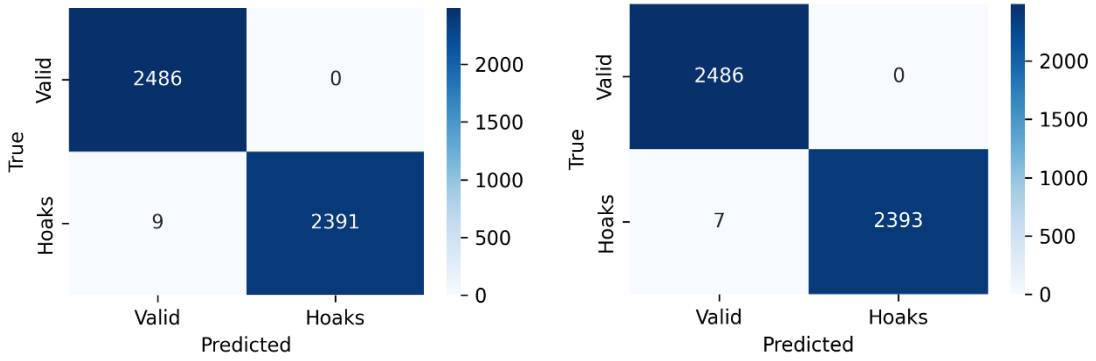
B. *Cross-Validation Performance*

To obtain a more robust estimate of generalization performance, both architectures were further evaluated using 10-fold stratified cross-validation. For each fold, the tokenizer was fitted only on the training subset and the model was trained from scratch, in line with standard practice in deep text classification [16]. The resulting accuracy, precision, recall, and F1-score on each validation fold were recorded and aggregated. Table 2 reports the mean and standard deviation of the metrics across the ten folds. The single-kernel 1D-CNN achieved an average accuracy of 99.898% ± 0.044, precision of 99.933% ± 0.076, recall of 99.858% ± 0.088, and F1-score of 99.896% ± 0.045. The proposed Multi-Kernel 1D-CNN slightly improves upon these results, obtaining an average accuracy of 99.918% ± 0.055, precision of 99.950% ± 0.058, recall of 99.883% ± 0.112, and F1-score of 99.917% ± 0.056.

TABLE 2. TEN-FOLD CROSS-VALIDATION RESULTS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Single-Kernel 1D-CNN	99.898 ± 0.044	99.933 ± 0.076	99.858 ± 0.088	99.896 ± 0.045
Multi-Kernel 1D-CNN (MK)	99.918 ± 0.055	99.950 ± 0.058	99.883 ± 0.112	99.917 ± 0.056

Although the absolute gains are modest (on the order of 0.02 percentage points), they are consistent across all four metrics and across folds, indicating that the multi-kernel design yields more stable and slightly more discriminative representations than the single-kernel baseline. This aligns with findings from multi-kernel CNN applications in other domains, where improvements of a few tenths of a percent are often considered meaningful on already strong baselines [18, 22].



(a) (b)

Figure. 8. Confusion Matrices for Single-Kernel CNN (a) and Multi-Kernel CNN (b)

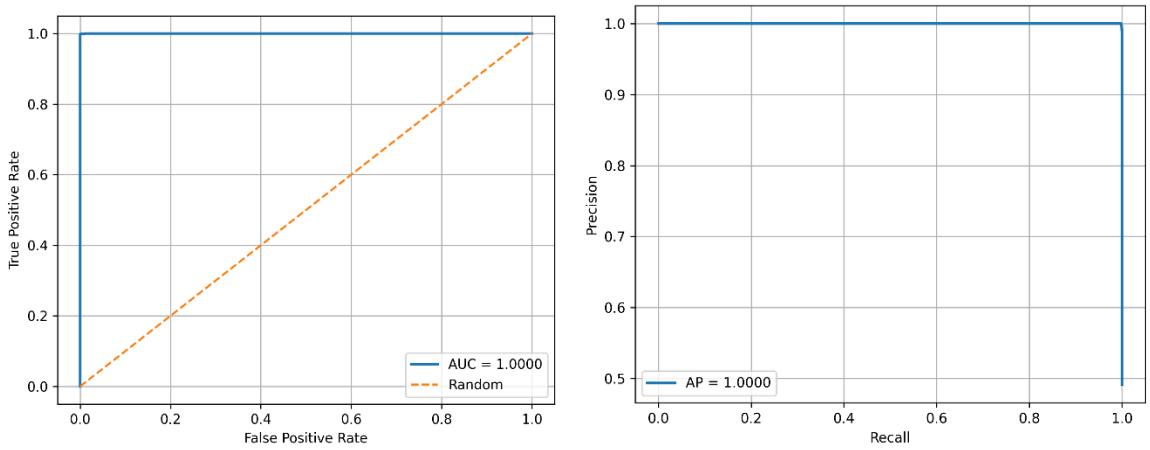


Figure. 9. ROC and Precision–Recall curves

Confusion matrices for both architectures on the hold-out test set (shown in Figure 8) further illustrate their behavior. In both cases, the number of misclassified samples false positives and false negatives is minimal compared to the total number of instances, with overall error rates well below 1%. The MK-CNN tends to reduce both types of errors slightly, leading to a higher F1-score. ROC and Precision–Recall curves (see Figure 9) show near-perfect separation, with curves closely hugging the upper-left and upper-right regions of the plot, respectively. This pattern is consistent with the extremely high F1 scores. It indicates that both models rank hoax and valid news very reliably across a range of decision thresholds, similar to results reported by recent deep-learning-based fake news detectors [7, 8, 24].

C. Comparative Analysis and Limitations

When compared with previous Indonesian hoax-detection studies, the proposed MK-CNN achieves competitive or superior performance. CNN-based approaches that operate solely on textual content typically report accuracies in the range of 95–98%

for Indonesian news datasets [23, 25], while hybrid IndoBERT-based models can reach similar high performance but often with higher computational cost and more complex deployment pipelines [26, 27]. In contrast, the MK-CNN used in this work attains an average accuracy above 99.9% while maintaining a relatively lightweight architecture that is straightforward to serve through a web API. The performance gain can be attributed to three factors: (1) a carefully designed preprocessing pipeline that removes label-leakage cues such as "hoax" and fact-check tags, (2) the multi-kernel convolution design that captures patterns at multiple n-gram scales, and (3) the multimodal treatment of OCR text from images and videos, which increases the available evidence for the classifier.

Despite these strong quantitative results, several limitations must be acknowledged. First, the dataset is built from curated news portals (Kompas.com) and a fact-checking site (TurnBackHoax.id). Although this setting is standard in the literature [23, 24], it may still be easier than detecting hoaxes in noisy, user-generated social-media content. Second, while the preprocessing steps were designed to mitigate label leakage, subtle source-specific cues or topic biases may remain, contributing to the very high scores. Evaluating the model on truly out-of-domain data such as tweets, WhatsApp forwards, or user comments would provide a more comprehensive picture of its robustness, as suggested by recent surveys on fake news detection [3-5]. Third, the current multimodal design reduces visual information to OCR text alone; integrating richer image features via joint text–image encoders [9, 10, 20] could further improve robustness, especially against adversarially manipulated images where text alone is not sufficient.

D. Deployment and Web-Based

The final phase of the CRISP-DM framework focuses on deployment, where the best-performing MK-CNN model is integrated into a web-based prototype for real-time hoax screening. The backend is implemented as a RESTful API that exposes endpoints for text classification and OCR-enhanced multimodal inference. When a user submits a news title, article body, and optional image or video, the server-side pipeline performs OCR (if media are present), applies the same preprocessing and tokenization steps used during training, and forwards the resulting sequence to the MK-CNN for prediction. This design ensures strict consistency between offline experiments and online inference, as recommended in earlier deployments of fake-news detection tools [24].

Figure 10 shows the user interface of the deployed web application. At the top, the interface presents a header and a brief description of the system, followed by an input panel where users can enter the news title and body, and optionally upload an image or video. The prediction panel on the right displays the MK-CNN's hoax probability, a categorical label (VALID or HOAX), and an indication of whether OCR was used in the current prediction. A threshold slider lets advanced users adjust the decision boundary to favor recall over precision, or vice versa.

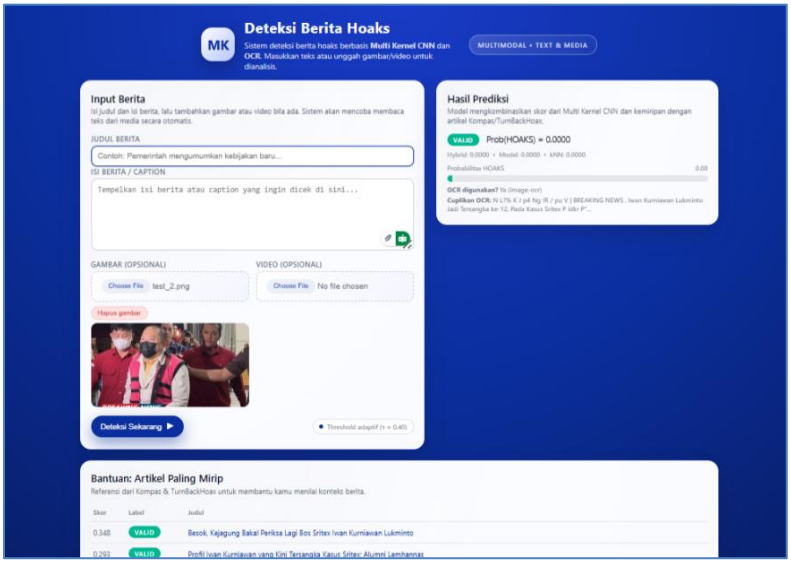


Figure. 10. Screenshot of the web-based hoax detection interface

The lower part of the interface provides decision-support information in the form of a ranked list of the most similar reference articles drawn from the Kompas.com and TurnBackHoax.id collections. Each row displays a similarity score, the predicted label of the reference article, and its title; clicking the title opens the original article in a new tab. This feature is inspired by earlier work on fake-news analysis systems that combine automatic classification with human-readable explanations or references [11, 24], and is intended to help journalists and lay users critically assess the model's prediction instead of accepting it blindly. Informal feedback from several trial users (students and peers) indicates that the combination of a clear probability

output, textual explanation, and reference articles makes the system intuitive and practically useful in everyday information checking.

Overall, the deployment results demonstrate that the proposed multimodal OCR + MK-CNN model is not only effective in offline evaluation, but also feasible for real-time use in a production-like environment. The web-based prototype operationalizes the complete CRISP-DM cycle from data understanding and modeling to deployment [21], providing a concrete example of how research on Indonesian hoax detection can be translated into a usable tool for the public.

V. CONCLUSION

This paper presented a multimodal framework for detecting Indonesian hoax news that combines Optical Character Recognition (OCR) with a Multi-Kernel one-dimensional Convolutional Neural Network (MK-CNN). Guided by the CRISP-DM process model, we constructed a balanced dataset of 24,439 news items from Kompas.com (valid news) and TurnBackHoax.id (verified hoaxes), designed a strict preprocessing pipeline to remove label-leakage cues and source-specific artifacts, and integrated OCR outputs from images and videos into a unified textual representation. The resulting sequences were modeled using both a single-kernel 1D-CNN baseline and the proposed MK-CNN architecture, enabling a systematic comparison under identical experimental settings.

Experimental results on an 80:20 hold-out split and 10-fold stratified a cross-validation demonstrate that both CNN-based models achieve very high performance, with average F1-scores above 99.8%. The proposed MK-CNN consistently outperforms the single-kernel baseline across accuracy, precision, recall, and F1-score, albeit with modest absolute gains, confirming that parallel convolutional filters with different kernel sizes can capture richer n-gram patterns without introducing substantial overfitting. Confusion matrices, ROC curves, and Precision–Recall curves further show that the model separates hoax and valid news almost perfectly across a wide range of decision thresholds. These results are competitive with, and in many cases superior to, previously reported Indonesian hoax-detection approaches based on CNN or Transformer architectures.

Beyond offline evaluation, the best-performing MK-CNN model was deployed in a web-based prototype that supports real-time hoax screening. The system accepts text, image, and video inputs, applies OCR when media are present, and returns a hoax probability and label together with a ranked list of similar reference articles from Kompas.com and TurnBackHoax.id. This deployment demonstrates that the proposed approach is not only effective in a controlled experimental setting but also practical as a decision-support tool for journalists, fact-checkers, and the general public. The end-to-end implementation—from data understanding and modeling to deployment illustrates how research on fake-news detection can be translated into a usable application in the Indonesian context.

Despite its strong performance, the proposed framework has several limitations. The current dataset is restricted to curated news portals and a fact-checking site, which may not fully reflect the noisier and more diverse content found on social media platforms. In addition, the multimodal component focuses on OCR-extracted text and does not yet exploit richer visual features such as objects, faces, or stylistic cues. As future work, we plan to (i) evaluate the model on out-of-domain data from social media and messaging applications, (ii) explore hybrid architectures that combine MK-CNN with pre-trained language models such as IndoBERT, and (iii) integrate joint text–image encoders to handle visually better manipulated hoaxes. User-centered studies on interpretability and probability calibration are also promising directions to ensure that automated hoax detectors can be safely and responsibly adopted in real-world settings.

REFERENCE

[1] E. Aïmeur, Z. Amri, and J. F. S. Ojo, "Fake news, disinformation and misinformation in social media: a review," Social Network Analysis and Mining, vol. 13, no. 1, 2023, doi: 10.1007/s13278-023-01028-5.

[2] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake News Detection on Social Media: A Data Mining Perspective," ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, 2017, doi: 10.48550/arXiv.1708.01967.

[3] A. Orhan, "Fake news detection on social media: the predictive role of university students' critical thinking dispositions and new media literacy," Smart Learning Environments, vol. 10, no. 1, 2023, doi: 10.1186/s40561-023-00248-8.

[4] M. M. S. Nasser et al., "A systematic review of multimodal fake news detection on social media using deep learning models," Results in Engineering, vol. 26, 2025, art. 104752, doi: 10.1016/j.rineng.2025.104752.

[5] I. Q. Abduljaleel and I. H. Ali, "Deep Learning and Fusion Mechanism-based Multimodal Fake News Detection Methodologies: A Review," Engineering, Technology & Applied Science Research, vol. 14, no. 4, pp. 15665–15675, 2024, doi: 10.48084/etasr.7907.

[6] A. S. Saini and P. S. Khatarkar, "A Review on Fake News Detection using Machine Learning," International Journal of Science and Research in Computer Science and Engineering and Information Technology, vol. 9, no. 2, 2023, doi: 10.24113/ijoscience.v9i2.511.

[7] R. K. Kaliyar, A. Goswami, P. Narang, and S. Sinha, "FNDNet: A Deep Convolutional Neural Network for Fake News Detection," Cognitive Systems Research, vol. 61, pp. 32–44, 2020, doi: 10.1016/j.cogsys.2019.12.005.

[8] R. K. Kaliyar, A. Goswami, P. Narang, and V. Chamola, "Understanding the Use and Abuse of Social Media: Generalized Fake News Detection with a Multichannel Deep Neural Network," IEEE Transactions on Computational Social Systems, vol. 11, no. 4, pp. 4878–4887, 2024, doi: 10.1109/TCSS.2022.3221811.

[9] P. K. Verma, P. Agrawal, V. Madaan, and R. Prodan, "MCred: multi-modal message credibility for fake news detection using BERT and CNN," Journal of Ambient Intelligence and Humanized Computing, vol. 14, pp. 10617–10629, 2023, doi: 10.1007/s12652-022-04338-2.

[10] A. Al-alshaqi et al., "A BERT-Based Multimodal Framework for Enhanced Fake News Detection Using Text and Image Data Fusion," Computers, vol. 14, no. 6, 2025, art. 237, doi: 10.3390/computers14060237.

[11] M. Patel and K. Surati, "MTLFND: Multimodal fake news detection using attention mechanism and transfer learning," Journal of Integrated Science and Technology, vol. 13, 2025, pp. 1–9, doi: 10.62110/sciencein.jist.2025.v13.1138.

[12] D. Mangal and D. K. Sharma, "Fake News Detection with Integration of Embedded Text Cues and Image Features," in Proc. 8th Int. Conf. on Reliability, Infocom Technologies and Optimization (ICRITO), 2020, pp. 68–72, doi: 10.1109/ICRITO48877.2020.9197817.

[13] A. B. Athira, A. Tiwari, S. D. M. Kumar, and A. M. Chacko, "Multimodal Data Fusion Framework For Fake News Detection," in INDICON 2022 – IEEE 19th India Council International Conference, 2022, doi: 10.1109/INDICON56171.2022.10039737.

[14] J. Memon, M. Sami, R. A. Khan, and M. Uddin, "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)," IEEE Access, vol. 8, pp. 142642–142668, 2020, doi: 10.1109/ACCESS.2020.3012542.

[15] N. H. Imam, V. G. Vassilakis, and D. Kolovos, "OCR post-correction for detecting adversarial text images," Journal of Information Security and Applications, vol. 66, 2022, art. 103170, doi: 10.1016/j.jisa.2022.103170.

[16] R. Juneja, G. Sikhwal, and J. Kumar, "Optical Character Recognition (OCR) Using Deep Learning and OpenCV," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 5, no. 3, pp. 553–558, 2019, doi: 10.32628/CSEIT195386.

[17] Y. Zhao, "Researches Advanced in Natural Scenes Text Detection Based on Deep Learning," Highlights in Science, Engineering and Technology, vol. 16, pp. 36–43, 2025, doi: 10.54097/hset.v16i.2500.

[18] Y. Kim, "Convolutional Neural Networks for Sentence Classification," in Proc. EMNLP 2014, pp. 1746–1751, 2014, doi: 10.3115/v1/D14-1181.

[19] H. He and Y. Yang, "The Text Classification Method Based on BiLSTM and Multi-Scale CNN," Computing and Communication, vol. 10, no. 1, 2022, doi: 10.54097/ypxxse31.

[20] Q. Hu, "Cross-Language News Text Classification Using a BERT-MCNN Model," Molecular & Cellular Biomechanics, vol. 20, no. 2, pp. 1–9, 2025, doi: 10.62617/mcb739.

[21] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," Procedia Computer Science, vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.

[22] M. K. N. Mursalim and A. Kurniawan, "Multi-kernel CNN block-based detection for COVID-19 with imbalance dataset," International Journal of Electrical and Computer Engineering, vol. 11, no. 3, pp. 2467–2476, 2021, doi: 10.11591/ijece.v11i3.pp2467-2476.

[23] M. A. Al Amin, D. Suhartono, and E. Budianto, "Indonesian Hoax News Detection Using One-Dimensional Convolutional Neural Network," Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI), vol. 14, no. 2, 2025, doi: 10.22146/jnteti.v14i2.19050.

[24] 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.

[25] R. Yoviananda and F. Fahrudin, "Implementation of Deep Learning to Detect Indonesian Hoax News with CNN Method," International Journal of Electrical and Electronic Engineering and Information Technology, vol. 4, no. 2, pp. 86–90, 2020, doi: 10.29138/ijeeit.v4i2.1525.

[26] M. Y. Ridho and E. Yulianti, "From Text to Truth: Leveraging IndoBERT and Machine Learning Models for Hoax Detection in Indonesian News," Jurnal Ilmiah Teknik Elektro Komputer dan Informatika, vol. 10, no. 3, pp. 544–555, 2024, doi: 10.26555/jiteki.v10i3.29450.

[27] D. Y. Yefferson, V. Lawijaya, and A. S. Girsang, "Hybrid model: IndoBERT and long short-term memory for detecting Indonesian hoax news," IAES International Journal of Artificial Intelligence, vol. 13, no. 2, pp. 1913–1924, 2024, doi: 10.11591/ijai.v13.i2.pp1913-1924.