Translation of article 10.22146/jnteti.v14i2.19050

Indonesian Hoax News Detection Using One-Dimensional Convolutional Neural Network

Muhammad Zuama Al Amin[1], Muhammad 'Ariful Furqon[2], Dwi Wijonarko[1]

[1] Program Studi Teknologi Informasi, Fakultas Ilmu Komputer, Universitas Jember, Jember, Jawa Timur 68121, Indonesia
[2] Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Jember, Jember, Jawa Timur 68121, Indonesia

[Submitted: 25 January 2025, Revised: 14 April 2025, Accepted: 28 May 2025]
Corresponding author: Muhammad 'Ariful Furqon (email: ariful.furqon@unej.ac.id)

ABSTRACT — The rapid advancement of information technology has enabled global information dissemination and led to a surge in hoax news, particularly in Indonesia. Hoax news poses a significant risk of spreading disinformation, potentially influencing public opinion, social stability, and security. Therefore, an effective technology-based solution is required to detect and identify hoaxes. This study aims to develop and optimize a one-dimensional convolutional neural network (1D-CNN) model to detect hoax news with high accuracy. The dataset comprised 12,151 articles, including 5,276 valid news items and 6,875 hoax news items, collected from reliable sources and anti-hoax platforms. The text preprocessing stages included data cleaning, case folding, punctuation removal, number removal, and stopword removal. The textual data were processed through tokenization and padding stages for model training preparation. The proposed 1D-CNN architecture integrated embedding, Conv1D, batch normalization, globalmaxpooling1d, dense, and dropout layers to enhance generalization capabilities and reduce the risk of overfitting. The model was trained using the Adam optimizer and its performance was evaluated using 10-fold cross-validation. Experimental results showed that the model achieved an average accuracy, precision, recall, and F1 score of 97.74%, 97.75%, 97.74%, and 97.73%, respectively. The developed model outperformed previous methods, namely the convolutional neural network–bidirectional long short-term memory (CNN-BiLSTM), gated recurrent unit (GRU), and conventional methods such as naïve Bayes or support vector machine (SVM), in terms of accuracy and training efficiency. This study demonstrates that the model has a reliable capability in identifying hoax news, both in terms of detection accuracy and performance consistency.

KEYWORDS — 1D-CNN, Hoax News Detection, Text Classification, Deep Learning.

I. INTRODUCTION

The advancement of technology has facilitated access to information and its dissemination, particularly through the Internet. Digital platforms such as social media and news websites have accelerated and expanded interpersonal interactions. However, this convenience has also opened significant opportunities for the spread of hoax information, which is often difficult to identify. According to the Kamus Besar Bahasa Indonesia (KBBI), a hoax is defined as false information deliberately disseminated to deceive or mislead [1]. Hoaxes are difficult to detect due to the overwhelming flow of information that hinders verification efforts. The primary media for hoax dissemination include blogs, news websites, and social media platforms [2].

Data from the Indonesian Ministry of Communication and Information Technology indicate that over 10,000 hoax cases have been reported in Indonesia over the past five years [3]. This phenomenon is further exacerbated by the low level of information literacy among the Indonesian public. This is evidenced by Indonesia's ranking of 71st out of 81 countries in the 2022 PISA survey on literacy [4]. Such conditions result in a tendency among the public to passively accept information without verification, thereby increasing the potential for hoax dissemination.

To address this challenge, an effective hoax detection system is needed to assist the public in identifying valid and false information [1]. Such systems can mitigate the spread of hoaxes by providing practical and accurate tools for information verification. Various approaches based on artificial intelligence, machine learning, and deep learning have been developed to tackle this issue. However, before implementing a deep learning or machine learning-based detection system, an appropriate model must be selected to ensure effective deployment [1].

The primary approach for hoax news detection is supervised learning. In this approach, models are trained on labeled data to recognize patterns within the text [5]. The primary approach for hoax news detection is supervised learning. In this approach, models are trained on labeled data to recognize patterns within the text [5]. Among the various classification methods, deep learning models are more prominent than conventional machine learning methods due to their ability to identify complex patterns in textual data.

One of the deep learning models that has demonstrated high accuracy in detecting hoax news is the one-dimensional convolutional neural network (1D-CNN). This model processes text as a 1D vector, enabling efficient recognition of important patterns in the data. Compared to traditional methods such as naïve Bayes and support vector machine (SVM) [1], the 1D-CNN exhibits significant advantages in both accuracy and generalization capability. However, due to its simple architecture, the 1D-CNN model is highly susceptible to overfitting when implemented on large datasets.

Prior studies have shown that the 1D-CNN model is highly effective in detecting hoax news written in Indonesian. The model is considered flexible as it can be combined with other techniques and is capable of achieving high levels of accuracy. Reference [1] reports that a 1D-CNN combined with batch normalization techniques was able to detect Indonesian hoax news with an accuracy of up to 97.9%. In addition, the 1D-

CNN model successfully detected hoaxes in Indonesian-language news articles with an accuracy of 99% [3]. In [6], the 1D-CNN model achieved an accuracy of 96.51% in detecting hoax news on social media platforms. These advantages highlight the 1D-CNN as a highly promising approach for practical implementation, especially amid the growing demand for intelligent solutions to counter the spread of hoaxes.

This study aims to optimize the 1D-CNN model by applying batch normalization and incorporating dropout layers to improve the model's generalization capability while reducing the risk of overfitting [7]. The primary objective of this research is to evaluate the performance of the 1D-CNN model in detecting hoax news based on textual data, particularly within the context of Indonesian-language news, and to achieve optimal performance from the 1D-CNN model. This study offers an innovative approach by optimizing the 1D-CNN architecture through a combination of dropout layers and batch normalization strategies, designed to address overfitting challenges in the implementation of deep learning models on large-scale datasets. This approach is expected to enhance the model's efficiency and accuracy in detecting linguistic and semantic patterns in hoax news data.

The main contribution of this research lies in the optimization of a deep learning-based hoax detection model trained using Indonesian-language news datasets. The optimization is performed by implementing a 1D-CNN model architecture enhanced with batch normalization and the addition of dropout layers, aimed at improving the model's ability to distinguish hoax news from valid news. This study also includes a comprehensive evaluation of the model's performance on textual data in a local context, providing new insights into the development of more specific and relevant fake news detection technologies in Indonesia.

The principal benefit of this study is the provision of a deep learning model as a practical solution for the rapid and accurate detection of hoax news. The developed model is expected not only to improve the effectiveness of false information detection but also to contribute to the advancement and optimization of deep learning technology for hoax news detection in the Indonesian language context.

II. HOAX NEWS

According to the KBBI, false news or hoaxes are defined as inaccurate information that is deliberately created and disseminated as if it were true. This type of information aims to deceive, mislead, or manipulate the recipients into believing in something that did not actually occur. Hoaxes are frequently spread through various media platforms, particularly in the current digital era where information dissemination can occur rapidly and without adequate verification [2].

One of the distinguishing features that can indicate whether news is a hoax is a provocative and exaggerated writing style. Hoaxes tend to use bombastic sentences designed to evoke strong emotional responses from readers [8], such as anger or fear. This contrasts with valid news, which typically adheres to a more formal and objective style of writing. These differences in sentence structure and writing style serve as key indicators in verifying the authenticity of news articles [2].

III. ONE-DIMENSIONAL CONVOLUTIONAL NEURAL NETWORK

A 1D-CNN is a neural network architecture specifically designed to process one-dimensional data, such as time-series signals, textual data, or other sequential forms of input. Unlike the two-dimensional convolutional neural network (2D-CNN) architecture commonly used for image processing, 1D-CNNs focus on linear data patterns and are widely applied in signal analysis, natural language processing, and pattern recognition tasks [9].

As illustrated in Figure 1, the 1D-CNN model is highly relevant for hoax news detection, as it can classify text into specific categories, such as hoax or valid news [5]. This model is designed to capture temporal or sequential patterns in textual data by converting it into numerical vectors, thereby enabling more effective text processing. Within a supervised learning framework, the model is trained using labeled datasets. During the training process, the network weights are optimized based on prediction errors relative to the given labels, allowing the model to accurately learn the characteristics of hoax news [10]. The detection process is conducted by analyzing sentence structures through convolutional layers that extract features from the sequence of words or tokens within the text. These layers are designed to detect local patterns such as word arrangements, phrase combinations, and inter-word relationships in specific contexts, which encompass both linguistic and semantic patterns useful for distinguishing fake news from genuine news. Additionally, the model leverages inter-feature correlations in the labeled dataset, where each data instance is associated with a specific news category (hoax or valid) [11]. These correlations help the model identify class-relevant patterns, thereby improving detection accuracy and generalization capability when encountering new data.

Previous studies have demonstrated that CNNs are highly effective in text analysis, including news classification and anomaly detection tasks [12]. The 1D-CNN model is capable of extracting linguistic features, such as word sequences, characteristic phrase patterns, and repetitive syntactic structures commonly found in hoax news texts. By employing 1D convolutions, the model can capture relationships between words or phrases without relying on the entire input sequence's order.

Compared to recurrent neural networks (RNNs), 1D-CNNs offer several advantages. RNNs typically require longer computation times due to their sequential nature and the use of mechanisms such as long short-term memory (LSTM) or gated recurrent units (GRU), which often increase model complexity. In contrast, 1D-CNNs are more computationally efficient as they extract local features through convolution operations, thereby reducing computational costs while maintaining high levels of accuracy [3].

IV. METHODOLOGY

This research methodology comprises systematic steps in the development of a hoax news detection model using a 1D-CNN architecture. The research stages began with data collection from relevant sources, followed by the text preprocessing stage, which included data cleaning, case folding, and stopword removal. The processed data were then divided into training and testing datasets for the purposes of model training and evaluation.

Subsequently, the data were converted into numerical vector representations that could be processed by the deep learning model. In the model development stage, the architecture of the 1D-CNN was selected and designed, including the determination of convolutional, pooling, and fully connected layers. In addition, hyperparameters such as the

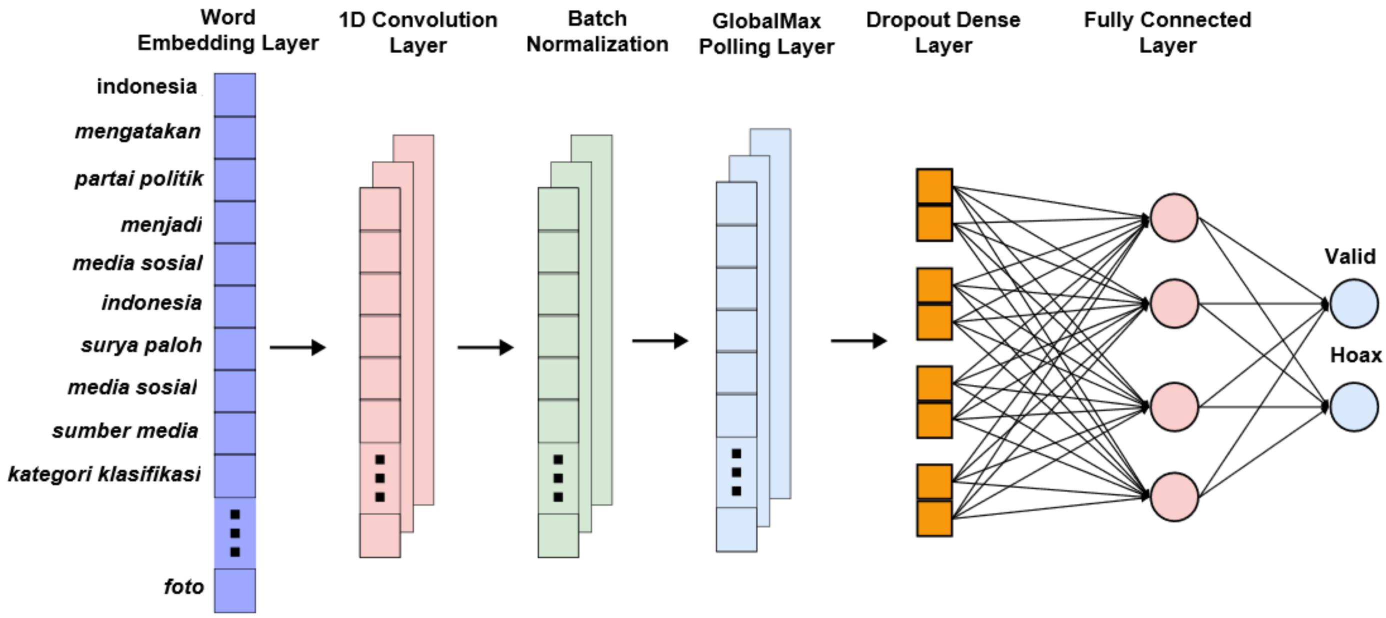

Figure 1. Model architecture of 1D-CNN.

number of epochs, batch size, learning rate, and optimizer were adjusted to optimize the model's performance.

Following the hyperparameter tuning, the model was trained using the training data. The training process includes periodic evaluations using performance metrics such as accuracy, precision, recall, and F1 score. Next, the model was tested using the testing data to assess its generalization ability and effectiveness in detecting hoax news. The evaluation results were then analyzed to interpret the model's performance and to provide recommendations for improvement in future studies.

A. DATA COLLECTION

Data collection in this study was based on a publicly available national news dataset that has been verified by experts in the field. The data consisted of Indonesian-language news articles, comprising both titles and content, where valid news articles are labeled as 0 and hoax news articles are labeled as 1. Valid news articles were obtained from open-source national news platforms that had undergone a labeling process, while hoax news articles were collected from the Turnbackhoax website [13], an anti-hoax community in Indonesia that specifically provides labels to differentiate hoax news from valid news.

These data serve as the primary input for model training and testing. Since the 1D-CNN model processes data in the form of word sequences, the news texts are first processed using tokenization techniques to convert each word into a numerical index representation. This representation allows the model to recognize relevant word sequence patterns in the news text, enabling effective classification.

B. TEXT PREPROCESSING

Text preprocessing is a crucial step to ensure that the data used are clean, relevant, and representative, in order to produce a model with optimal performance [14]. The first step in text preprocessing is data cleaning, which includes handling missing values by removing rows containing such values. The data cleaning process also involves removing duplicate data to prevent bias during model training and concludes with merging the cleaned data.

The next step is case folding, which converts all text to lowercase to ensure consistency so that words with different capitalizations are recognized as the same entity. This is followed by the removal of symbols and special characters, such as punctuation marks, numbers, or emojis that are irrelevant to the analytical context, resulting in cleaner and more structured data [14]. In addition, stopwords such as "dan" (and), "atau" (or), and "yang" (which), which do not contribute significantly to semantic analysis, are removed to reduce data dimensionality and enhance focus on the more important words [14]. Through this preprocessing, the data become more suitable for analysis and model training, thereby improving the accuracy of the final outcomes.

C. MODEL DEVELOPMENT AND TRAINING

At this stage, the development process of the 1D-CNN model began. The first step involved tokenizing the data. Here, the text prepared for training was converted into tokens or word units arranged according to word indices. This process produced numerical data representations, which were then converted into sequences so that they could be processed by the model. The sequential data were then transformed into a categorical format for classification purposes and subsequently divided into training and validation datasets. The next step was to define the hyperparameters and design the model architecture to ensure that the model could process the data efficiently and detect relevant patterns that support the classification objective.

Hyperparameters are values determined prior to training to control the architecture and learning process, such as embeddings and filters. The 1D-CNN model was built with an embedding layer to convert words into vectors, followed by a Conv1D layer to detect patterns in the sequences. Then, GlobalMaxPooling1D was used to reduce dimensionality, followed by a dense layer [1] and a dropout layer to prevent overfitting [15], and the model was trained for a total of 10 epochs [1]. The model was trained using these hyperparameters to optimize its performance.

D. MODEL EVALUATION

After the model was successfully built and trained, it was evaluated using the 10-fold cross-validation method to comprehensively assess its performance. This technique was chosen because it provided a more stable performance estimate compared to single-test methods and reduced the risk of overfitting by testing the model across multiple data subsets [16]. In each iteration, the dataset was divided into ten subsets, with nine subsets used for training and one subset used for testing. This process was repeated until all subsets had been used as test data. The model's prediction results were compared with the original labels to calculate evaluation metrics such as accuracy, precision, recall, and F1 score [17]. These metrics provided a comprehensive overview of the model's capabilities, from its accuracy in making predictions to its sensitivity in detecting hoax content. This evaluation was crucial not only for assessing the model's generalization ability on new data but also for identifying potential improvements in the model's architecture or training process. If any performance imbalance was observed in certain categories, adjustments were made, such as adding more data to underrepresented categories or optimizing the model's hyperparameters. This step ensured that the resulting model exhibited reliable performance and was ready for deployment in real-world scenarios.

V. RESULTS AND DISCUSSION

This discussion section is divided into four main parts. The first part describes the dataset collection process, including data sources and selection criteria. The second part discusses the results of text preprocessing, such as data cleaning, case folding, and stopword removal. The third part explains the model training and architecture, including hyperparameter selection and training outcomes. The fourth part discusses the model evaluation results using metrics such as accuracy, precision, recall, and F1 score to assess the effectiveness of hoax news detection.

A. DATA COLLECTION RESULTS

Data collection in this study refers to open-source national news datasets as well as publicly available datasets verified by domain experts. A total of 12,151 news articles were used, collected through two main stages. This process was designed to ensure that the dataset used was representative, high in quality, and supported the research objective, namely to build a high-performance and high-accuracy hoax news detection model. Additionally, the use of a large dataset was expected to enhance the model's generalization capability across various types of news.

In the initial stage, 1,246 Indonesian-language news articles were obtained from a previous study [18]. This dataset consisted of 553 valid news articles and 693 hoax news articles. A selection and verification process were conducted to ensure data quality, including checks for validity, relevance, and balance between valid and hoax news. This initial dataset served as the basis for preliminary experiments in developing the hoax news detection model.

In the second stage, the dataset was significantly expanded by adding 10,905 additional data points sourced from prior research and obtained through scraping various news portals, as described in the previous study [13]. The scraping process was carried out systematically by referring to two main categories of sources: trusted news sites and hoax news portals. This dataset expansion aimed to increase the diversity and coverage of the data, thereby strengthening the model's capability to detect fake news more accurately.

Valid news articles were collected from prominent Indonesian news portals such as Kompas and CNN, which have a strong reputation for delivering factual information. In contrast, hoax news articles were obtained from Turnbackhoax, a platform that specifically documents and verifies various false information circulating in society. The final collected dataset consisted of 5,276 valid news articles labeled as 0 and 6,875 hoax news articles labeled as 1, ensuring a good data balance for model training.

B. TEXT PREPROCESSING

The text preprocessing phase began with a data cleaning process, which involved removing incomplete or missing values. The dataset sourced from Kompas contained 27 unknown values; therefore, rows with missing values in the relevant columns were removed. After this process, the remaining data from Kompas totaled 4,723 rows.

The dataset from Turnbackhoax also had similar issues, with some rows containing incomplete values in certain columns. Rows lacking complete data were removed to improve dataset quality. After cleaning, the Turnbackhoax dataset consisted of 6,592 rows with various labels: *Misinformasi* (Misinformation), *Hoaks* (Hoax), *Benar* (True), and Salah (False). A filtering process was then performed on the Turnbackhoax dataset to ensure label consistency. Rows labeled as *Disinformasi* (Disinformation), *Salah*, and *Misinformasi* were retained, while those labeled *Benar* were removed so that the dataset only contained hoax-labeled news. After this stage, the total number of cleaned and usable data entries was 6,182 rows.

For the dataset obtained from the previous study [18], label normalization was performed: hoax-labeled data were relabeled as 1, and valid-labeled data were relabeled as 0. This dataset included 553 valid news articles and 693 hoax news articles. Once all datasets were confirmed to be clean, they were merged. However, the merged dataset was not yet fully ready for model training and thus required further preprocessing before being used as training data. The final dataset consisted of 12,151 news articles, including 6,874 hoax news articles (56.5%) and 5,276 valid news articles (43.5%). This distribution reflects a relatively balanced proportion. The dataset composition is presented in Table I.

Table I presents the cleaned and merged data, which still contains inconsistent text elements, such as capital letters, punctuation marks, and several unnecessary symbols. The next step was case folding, which involved converting all text to lowercase to ensure consistency, so that words like "Indonesia" and "indonesia" would not be treated differently by the model. In addition, the text was cleaned of non-alphabetical symbols that could interfere with the pattern recognition process in the data. After case folding, stopword removal was carried out to eliminate common words such as "*dan*" (and), "*atau*" (or), and "*yang*" (which), which did not contribute significantly to the analysis. This process utilized the Natural Language Toolkit (NLTK) and Sastrawi libraries, resulting in text that retained only essential words to help the model better understand the context [19].

The preprocessed dataset is shown in Table II, which demonstrates significant changes in the data structure, ensuring that the text is free from irrelevant elements such as symbols,

TABLE I
DATASETS THAT HAVE GONE THROUGH THE DATA CLEANING PHASE

Label	Text
1	*Anies Baswedan Dekat Dengan Aliran Krsiten Se...*
1	*Hakim Wahyu Iman Santoso Alami Kecelakaan Tun...*
1	*GAMBAR MEGAWATI DAN PUAN BERMAIN SLOT Nenek l...*
…	...
0	*Efek Ekor Jas Pencalonan Anies, Elektabilitas ...*
0	*Survei Litbang Kompas: PDI-P, Gerindra, dan Go...*

TABLE II
DATASETS THAT HAVE GONE THROUGH STOPWORD DELETION PHASE

Label	Text
1	*anies baswedan dekat aliran krsiten sesat alph...*
1	*hakim wahyu iman santoso alami kecelakaan tung...*
1	*gambar megawati puan bermain slot nenek lampir...*
1	*jonathan latumahina seorang nasrani penyusup o...*
0	*kemenangan telak anies sandiaga pilkada dki...*
...	...
0	*pesawat hercules milik tni au pesawat hercules chs milik tni angkatan udara...*

numbers, or common words that do not contribute meaningfully to the analysis. To provide further visualization, a word cloud was generated and is displayed in Figure 2. These steps ensured that the data met the quality standards required for further analysis and deep learning model development.

In Table II, stopwords such as "*dan*," "atau," "*dengan*," and other common words were removed from the text attribute using the Sastrawi library. The purpose of stopword removal was to eliminate words with minimal semantic value, leaving only relevant keywords. This process aimed to improve the quality of the "Text" column for subsequent analysis, such as hoax detection. Figure 2 presents a word cloud that visualizes the most frequently occurring words in the dataset after preprocessing, providing initial insights into common patterns, themes, and dominant terms in the data [20].

C. MODEL TRAINING

Following the text underwent preprocessing, the data were ready to be used in the model training phase, specifically for hoax detection classification. This process began by converting the text into a categorical format suitable for the model. The textual data were taken from the dataset and processed further through tokenization, padding, and other transformations to prepare them for deep learning processing. The labels indicating the news category (e.g., "hoax" and "valid") were then converted into a binary format using one-hot encoding. This format converted the textual labels into representations such as [1, 0] for valid news and [0, 1] for hoax news. Examples of this process are presented in Table III, which lists several news text samples along with their corresponding category labels.

The tokenization process was a crucial step in text data processing. Text was converted into sequences of tokens representing words or subwords from a predefined vocabulary. During this process, words not found in the vocabulary (out-of-vocabulary/OOV) were replaced with a special token designed to handle such cases. Tokenization was conducted using the built-in tokenizer from the TensorFlow Keras library, which provides flexibility in configuring parameters such as *num_words* (the maximum number of words considered in the

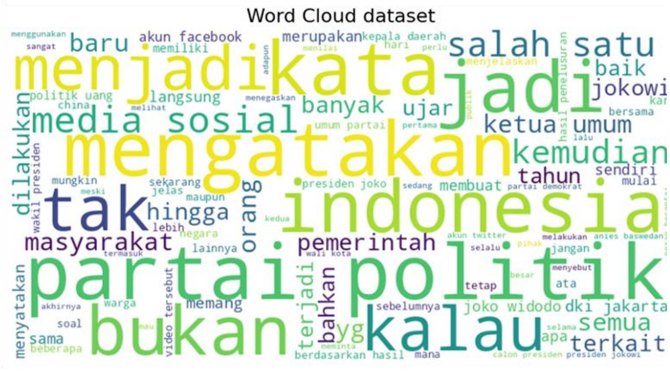

Figure 2. Word cloud dataset.

TABLE III
DATASET AFTER CONVERTED INTO CATEGORICAL

Text	Label	Remarks
luhut curigai motif politik balik penolakan wni wuhan natuna adanya penolakan masyarakat kabupaten natuna kepulauan riau warga negara	[True False]	The text is an example of valid news; therefore, the label is True False.
.....		
sepatu matamata perang dunia ii sepatu dipakai matamata sekutu selama perang dunia ii menyesatkan jerman mengarahkan kearah berlawanan	[False True]	The text is an example of hoax news; therefore, the label is False True.

vocabulary) and *oov_token* for identifying unknown words. This process ensured that all text could be converted into a numerical representation that the model could consistently process, including data with infrequent words.

Following tokenization, padding was applied using the pad_sequences function. Padding aimed to standardize the length of each text sequence so that the model could process inputs of consistent size. Texts shorter than the maximum length were padded with zeros, while longer texts were truncated according to the specified *trunc_type* parameter, such as "pre" or "post." This process was critical to ensure the model received uniform input, facilitating training and improving efficiency.

After tokenization and padding, the dataset was split into two parts: 80% for training and 20% for testing. The training data were used to train the model, while the testing data were used to evaluate the model's performance during training. This evaluation helped identify the best-performing model based on metrics such as accuracy and loss. By partitioning the dataset, the model could be tested on previously unseen data to ensure good generalization.

The final stage of this process involved building the 1D-CNN model architecture, which was designed to detect hoax news by identifying distinctive patterns in news texts [9]. The model utilized convolutional layers to extract significant features from the text data, which were subsequently passed to dense layers for final decision-making. Hyperparameters such as embedding size, number of filters, kernel size, activation function, learning rate, dropout rate, and loss function were adjusted to optimize the model's performance and prevent overfitting.

p-ISSN 2301–4156 | e-ISSN 2460–5719

The model began with an embedding layer, which functioned to convert text into fixed-dimensional vector representations [21]. This representation enabled the model to process textual data in a form it could understand while preserving contextual meaning. Following this, a Conv1D layer was applied to extract essential features from the token sequences, allowing the model to recognize patterns in the text data and to understand inter-word relationships within specific contexts.

To enhance training stability and efficiency, a batch normalization layer was added to normalize the output of the preceding layer. This process helped reduce internal covariate shifts during training, accelerate convergence, and improve the model's overall performance [22]. Subsequently, a GlobalMaxPooling1D layer was applied to reduce data dimensionality by selecting the maximum value across the sequence length as the most dominant feature [23] This layer not only reduced data complexity but also retained key information from the most relevant features in detecting patterns. This combination ensured that the model could capture more informative and stable data representations, thereby improving classification accuracy for hoax news.

The final layer consisted of a dense layer with two output units, utilizing a softmax activation function to classify texts into two categories: hoax or valid. To improve the model's generalization ability and mitigate overfitting, a dropout layer with a rate of 0.5 was added. This technique randomly deactivated a portion of neurons during training, preventing the model from relying too heavily on specific neurons. As a result, the dropout layer enabled the model to learn more generalized patterns, reduced the risk of overfitting, and enhanced its performance on unseen data. The addition of dropout was expected to improve model robustness and reliability under various data conditions [24].

The training process of the 1D-CNN model for hoax news detection was carried out using the preprocessed dataset. The training was conducted over ten epochs [1]. The results of the training process are illustrated in Figure 3, which presents performance metrics such as accuracy and loss values for both training and validation datasets across each epoch. The graph provides insight into the model's convergence behavior and its performance in detecting hoax news based on the utilized dataset.

As shown in Figure 3, the model achieved a training accuracy of 88.41% and a validation accuracy of 97.20% in the first epoch, indicating good initial capability in learning dataset patterns. The training accuracy continued to increase up to 98.16%, while the validation accuracy remained stable at 97.98%. A reduction in training loss from 0.2533 to 0.0408 indicated an effective learning process, although fluctuations in validation loss from 0.3171 to 0.1134 suggested a slight potential for overfitting. With an average duration of 9 seconds per epoch, the 1D-CNN model proved to be efficient and demonstrated good generalization capability.

The training results also showed a better and more balanced performance in a configuration using the 1D-CNN architecture with batch normalization but without the dropout layer [1] achieving 100% training accuracy. The replication of these results is illustrated in Figure 4.

Figure 4 shows that the model reached 100% training accuracy and 98% validation accuracy. The gap between training and validation accuracy, as well as the difference between training loss and validation loss, was relatively small. However, the perfect training accuracy may indicate a risk of overfitting, suggesting that the model's performance on testing data might not be optimal [24].

This outcome differs from the previous training that incorporated a dropout layer, in which training accuracy was 98.16% and validation accuracy remained stable at 97.98%. This comparison demonstrates that the use of dropout can help reduce the risk of overfitting. Therefore, further evaluation or the application of additional regularization techniques—such as dropout or model complexity reduction—may be necessary to enhance the model's generalization performance.

D. MODEL EVALUATION

The model was evaluated using the 10-fold cross-validation method to measure performance consistency across various data subsets [1]. In this method, the dataset was divided into 10 balanced folds, with each fold serving alternately as validation data, while the remaining 9 folds were used for training. This approach effectively reduced bias, improved evaluation reliability, and provided a more accurate representation of the model's capability to handle new data.

The performance results were analyzed using a confusion matrix to compute key metrics, including accuracy, precision, recall, and F1 score. Accuracy measured the proportion of total correct predictions, while precision calculated the proportion of relevant positive predictions. The recall was used to evaluate the model's ability to identify all actual positive samples, and the F1 score represented a harmonic metric that balances precision and recall. This analysis provided in-depth insight into the model's strengths and weaknesses in detecting hoax news with a high degree of validity. Based on the confusion matrix, these metrics were calculated using (1) through (4).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1\ Score = \frac{2 \cdot Precision \cdot Recall}{Precision+Recall}. \tag{4}$$

In addition to the confusion matrix, the evaluation results were compared with previous studies to assess the advantages or performance improvements of the proposed model. This analysis provided a comprehensive view of the model's effectiveness in detecting hoax news and highlighted the relevance of the proposed method in relation to existing approaches. The model evaluation results are presented in Table IV.

This comparative approach was important to ensure that the developed model not only performed well in theory but also demonstrated tangible superiority compared to other methods previously applied. For example, earlier studies employing naïve Bayes or SVM approaches may have achieved acceptable accuracy levels but often faced limitations in handling the high complexity of textual data. By employing a 1D-CNN model in this study, there was greater potential for the model to capture deeper patterns in the text, especially in detecting fake or hoax news, which often contains irregular sentence structures.

The comparative analysis with prior studies focused on evaluation metrics such as accuracy, precision, and recall to objectively assess the model's performance in classifying hoax news. This approach enabled a measurable and relevant evaluation directly aligned with the study's primary

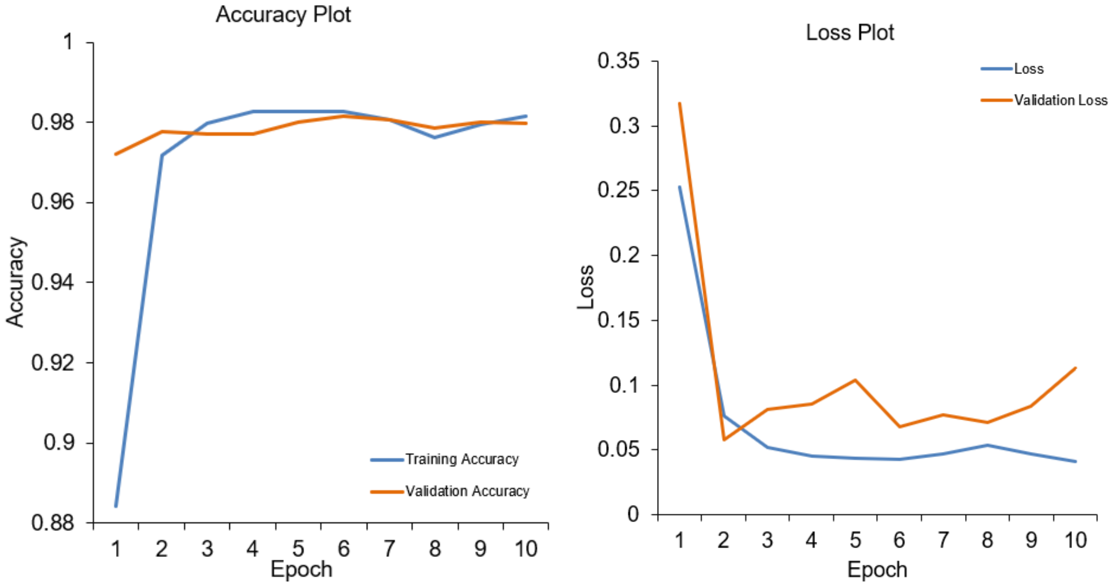

Figure 3. Graph of model training results.

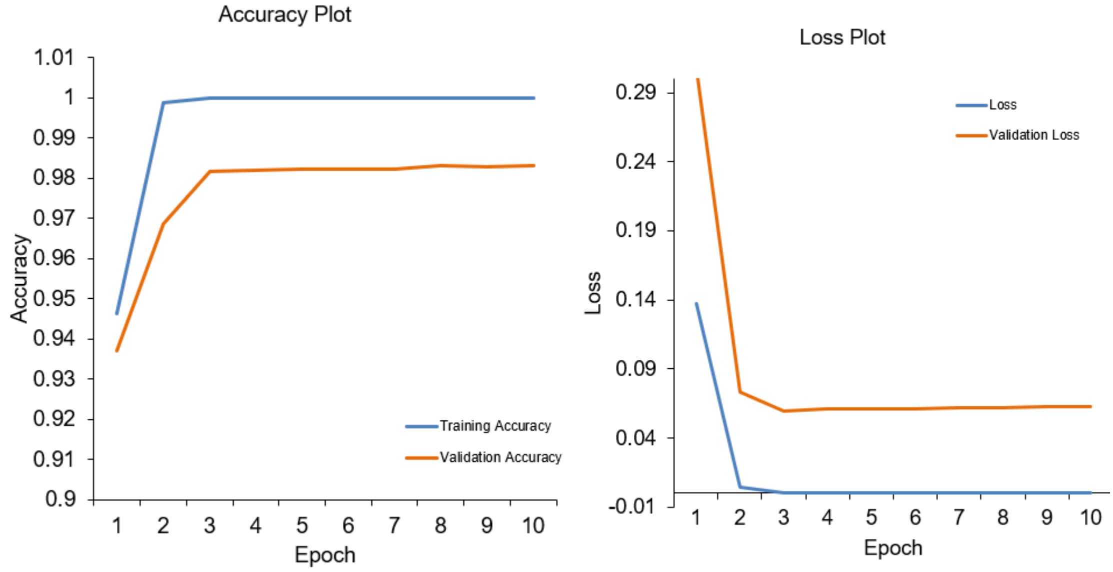

Figure 4. Graph of training results using the proposed model [1].

objective—enhancing the performance of text-based hoax news detection models. This is attributed to the 1D-CNN structure being specifically designed to efficiently process sequential data through the use of convolutional filters that can directly capture local features from the input text.

Furthermore, this study also evaluated the model's capability in handling various types of fake news. One of the main challenges in hoax detection is the variation in writing styles used by hoax creators to obscure false information. In this regard, the 1D-CNN approach exhibited advantages in recognizing specific patterns at both the character and word levels. This enabled the model to detect hoaxes with writing patterns similar to the training data, even when different vocabularies were used.

In previous studies, one commonly faced challenge was the low model performance when applied to data distributions that differed from the training data. To address this issue, the model in this study was designed using regularization techniques such as dropout and batch normalization to reduce the risk of overfitting.

TABLE IV
MODEL EVALUATION RESULTS

Metric	Fold 1	Fold 2	…	Fold 10	Average
Accuracy	0.9737	0.9786	…	0.9712	0.9774 (± 0.0040)
Precision	0.9740	0.9788	…	0.9712	0.9775 (± 0.0039)
Recall	0.9737	0.9786	…	0.9712	0.9774 (± 0.0040)
F1 score	0.9736	0.9786	…	0.9712	0.9773 (± 0.0040)

The evaluation results presented in Table IV indicate that the 1D-CNN model not only achieved higher accuracy but also demonstrated consistent performance across other metrics such as precision and recall compared to previous studies. In the context of fake news detection, high precision indicated the model's capability to minimize false positives—valid news articles misclassified as hoaxes. Conversely, high recall indicated the model's ability to detect most hoax news in the dataset, thereby reducing the risk of missing misleading information. By comparing the results of this study with other approaches applied in various contexts, it can be concluded that

TABLE V
TABLE OF COMPARISON WITH PRIOR RESEARCH

Model	Accuracy (%)	Precision (%)	Recall (%)
1D-CNN+ Batch Normalization +Dropout	97.74	97.75	97.74
BI-LSTM + Dropout [1]	96.60	96.56	96.54
Hybrid CNN-SVM [11]	95.99	-	-
Naïve Bayes [1]	86.20	86.95	86.20
Hybrid CNN-BI LSTM [6]	95.94	-	-
GRU [14]	90.00	90.00	87.00

the proposed 1D-CNN model has strong potential as an effective solution for automatic hoax news detection. This analysis not only contributes to the academic literature but also opens opportunities for further development, including deployment on broader platforms such as social media.

The average evaluation results showed an accuracy of 97.74% with a standard deviation of ±0.0040. The precision, recall, and F1 score values also remained consistently high, as presented in Table IV, demonstrating the model's strong generalization capability. Moreover, the model's performance in this study surpassed previous findings [6], which utilized a 1D-CNN architecture with dropout techniques as well as a hybrid 1D-CNN-BiLSTM model for hoax news detection. In that study, the 1D-CNN model achieved an accuracy of 96%. Thus, the approach proposed in this research not only succeeded in overcoming the risk of overfitting but also showed improved accuracy in detecting hoax news. A comparison of hoax news detection model performance with prior research is shown in Table V.

Based on Table V, it can be observed that the proposed model—1D-CNN + batch normalization + dropout—outperformed other models in the context of hoax news detection. This model demonstrated the highest performance, as reflected by the highest accuracy achieved.

The superiority of this model can be attributed to the 1D-CNN's ability to capture deep linguistic patterns in textual data, combined with the use of batch normalization and dropout techniques, which are effective in reducing overfitting [24]. Additionally, this approach leveraged the computational efficiency of the 1D-CNN, enabling faster training compared to hybrid CNN-BiLSTM models. Overall, this study demonstrated that the 1D-CNN + batch normalization + dropout approach not only excels in accuracy but also provides a simpler, more efficient, and effective solution for hoax news detection. These findings contribute significantly to the development of NLP-based hoax detection systems that can be practically implemented in the digital era, while also enhancing public information literacy. By adapting this model to the Indonesian language context, this research also opens potential applications in various languages and fields, providing a strong foundation for further studies in timely and accurate hoax detection.

VI. CONCLUSION

The 1D-CNN model demonstrated excellent performance in detecting and classifying hoax news, achieving an average accuracy of 0.9774, precision of 0.9775, recall of 0.9774, and an F1 score of 0.9773. This evaluation was conducted using 10-fold cross-validation on a dataset consisting of 12,151 news articles, with a distribution of 56% hoax news and 44% valid news. These results reflect the model's capability not only to detect the presence of hoax news but also to consistently and accurately classify news into hoax and valid categories. The detection process was carried out by identifying linguistic and semantic patterns through the 1D-CNN layers, which were capable of extracting features such as word sequences and phrase combinations to produce accurate predictions. The correlation between detection and classification performance was found to be strong, with the model's accuracy in detecting hoax news closely tied to its efficiency in recognizing the distinctive patterns of each news category. For future research, it is recommended to integrate stemming and lemmatization techniques in the preprocessing stage to enhance text representation. Additionally, consideration of dataset distribution across news categories—such as politics, technology, entertainment, education, and sports—is important to ensure that the model can capture broader contextual patterns and maintain balanced performance across categories. Further optimization of the model is expected to improve its performance. The implementation of the model in the form of an Android application or web-based platform for hoax detection aims not only to maintain the high model performance but also to ensure practical integration and usability for the general public.

AUTHORS' CONTRIBUTIONS

Conceptualization, Muhammad Zuama Al Amin and Muhammad 'Ariful Furqon; methodology, Muhammad Zuama Al Amin; software, Muhammad Zuama Al Amin; validation, Muhammad Zuama Al Amin, Muhammad 'Ariful Furqon, and Dwi Wijonarko; formal analysis, Muhammad Zuama Al Amin; investigation, Muhammad Zuama Al Amin; resources, Muhammad Zuama Al Amin; data curation, Muhammad Zuama Al Amin; writing-original drafting, Muhammad Zuama Al Amin; writing-reviewing and editing, Muhammad Zuama Al Amin; visualization, Muhammad Zuama Al Amin; supervision, Muhammad Zuama Al Amin; project administration, Muhammad Zuama Al Amin; funding acquisition, Muhammad 'Ariful Furqon.

ACKNOWLEDGMENT

The authors extend sincere gratitude to all parties who have provided support in the preparation of this research. In particular, the authors express appreciation to the Faculty of Computer Science, University of Jember, for the facilities and support provided throughout the research period. The authors are also thankful to colleagues and supervisors for their valuable input that contributed to the preparation of this manuscript. In addition, heartfelt appreciation is conveyed to our families and friends for their unwavering prayers and motivation. Lastly, the authors thank all individuals who, directly or indirectly, have contributed to this research.

REFERENCES

[1] B.P. Nayoga, R. Adipradana, R. Suryadi, and D. Suhartono, "Hoax analyzer for Indonesian news using deep learning models," *Procedia Comput. Sci.*, vol. 179, pp. 704–712, Feb. 2021, doi: 10.1016/j.procs.2021.01.059.

[2] Mukhlizar and F. Yuliani, "Hoax dan tanggung jawab sosial pengguna sosial media," *Kaganga, J. Pendidik. Sej. Ris. Sos. Hum.*, vol. 6, no. 2, pp. 461–466, Jul.-Dec. 2023, doi: 10.31539/kaganga.v6i2.7695.

[3] T.C. Praha, Widodo, and Murien Nugraheni, “Indonesian fake news classification using transfer learning in CNN and LSTM,” *JOIV, Int. J. Inform. Visualization,* vol. 8, no. 3, pp. 1213–1220, Sep. 2024, doi: 10.62527/joiv.8.2.2126.

[4] D. Amelia *et al.,* “Variabel yang memengaruhi kemampuan literasi membaca siswa Indonesia: Analisis berdasarkan pendekatan MARS,” *J. Pendidik. Kebud.,* vol. 9, no. 2, pp. 205–217, Dec. 2024, doi: 10.24832/jpnk.v9i2.4966.

[5] J.C.S. Reis *et al.,* “Supervised learning for fake news detection,” *IEEE Intell. Syst.*, vol. 34, no. 2, pp. 76–81, Mar./Apr. 2019, doi: 10.1109/MIS.2019.2899143.

[6] A.D. Safira and E.B. Setiawan, “Hoax detection in social media using bidirectional long short-term memory (Bi-LSTM) and 1 dimensional-convolutional neural network (1D-CNN) methods,” in *2023 11th Int. Conf. Inf. Commun. Technol. (ICoICT)*, 2023, pp. 355–360, doi: 10.1109/ICoICT58202.2023.10262528.

[7] M.F. Mridha *et al.,* “A comprehensive review on fake news detection with deep learning,” *IEEE Access*, vol. 9, pp. 156151–156170, Nov. 2021, doi: 10.1109/ACCESS.2021.3129329.

[8] M.G. Ayu, “Peneliti manfaatkan AI untuk deteksi berita palsu COVID-19.” Cloud Computing Indonesia. Access date: 20-Nov-2024. [Online]. Available: https://www.cloudcomputing.id/berita/peneliti-manfaatkan-ai-dalam-mendeteksi-berita-palsu

[9] S. Kiranyaz *et al.,* “1D convolutional neural networks and applications: A survey,” *Mech. Syst. Signal Process.*, vol. 151, pp. 1–21, Apr. 2021, doi: 10.1016/j.ymssp.2020.107398.

[10] R. Oshikawa, J. Qian, and W. Y. Wang, “A survey on natural language processing for fake news detection,” in *Proc. 12th Lang. Resour. Eval. Conf.,* 2020, pp. 6086–6093.

[11] M. Benedict and E.B. Setiawan, “Hoax detection on social media with convolutional neural network (CNN) and support vector machine (SVM),” in *2023 11th Int. Conf. Inf. Commun. Technol. (ICoICT)*, 2023, pp. 361–366, doi: 10.1109/ICoICT58202.2023.10262433.

[12] X. Zhang, J. Zhao, and Y. LeCun, “Character-level convolutional networks for text classification,” in *NIPS'15, Proc. 29th Int. Conf. Neural Inf. Process. Syst.,* 2015, pp. 649–657.

[13] O.C.R. Rachmawati and Z.M.E. Darmawan, “The comparison of deep learning models for Indonesian political hoax news detection,” *CommIT (Commun. Inf. Technol.) J.,* vol. 18, no. 2, pp. 123–135, Oct. 2024, doi: 10.21512/commit.v18i2.10929.

[14] K.S. Nugroho, “Dasar text preprocessing dengan Python.” Medium. Access date: 20-Oct-2024. [Online]. Available: https://ksnugroho.medium.com/dasar-text-preprocessing-dengan-python-a4fa52608ffe

[15] L. Hu, S. Wei, Z. Zhao, and B. Wu, “Deep learning for fake news detection: A comprehensive survey,” *AI Open*, vol. 3, pp. 133–155, Oct. 2022, doi: 10.1016/j.aiopen.2022.09.001.

[16] J. Brownlee, “A gentle introduction to k-fold cross-validation.” Access date: 30-Oct-2024. [Online]. Available: https://machinelearningmastery.com/k-fold-cross-validation/

[17] S. Narkhede, “Understanding confusion matrix.” Towards Datascience. Access date: 20-Oct-2024. [Online]. Available: https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

[18] S.F.N. Azizah, H.D. Cahyono, S.W. Sihwi, and W. Widiarto, “Performance analysis of transformer based models (BERT, ALBERT and RoBERTa) in fake news detection,” in *2023 6th Int. Conf. Inf. Commun. Technol. (ICOIACT),* 2023, pp. 425–430, doi: 10.1109/ICOIACT59844.2023.10455849.

[19] M.A. Rosid *et al.,* “Improving text preprocessing for student complaint document classification using Sastrawi,” in *Int. Conf. Eng. Technol. Appl. Sci. (ICETsAS),* 2020, pp. 1–6, doi: 10.1088/1757-899X/874/1/012017.

[20] M.G. Pradana, “Penggunaan fitur wordcloud dan document term matrix dalam text mining,” *J. Ilm. Inform.,* vol. 8, no. 1, pp. 38-43, Mar. 2020, doi: 10.33884/jif.v8i01.1838.

[21] R. Egger, “Text representations and word embeddings,” in *Applied Data Science in Tourism*. Cham, Switzerland: Springer, 2022, pp. 335–361.

[22] C. Garbin, X. Zhu, and O. Marques, “Dropout vs. batch normalization: An empirical study of their impact to deep learning,” *Multimed. Tools Appl.,* vol. 79, no. 19–20, pp. 12777–12815, May 2020, doi: 10.1007/s11042-019-08453-9.

[23] H. Kim and Y.S. Jeong, “Sentiment classification using convolutional neural networks,” *Appl. Sci.,* vol. 9, no. 11, pp. 1–14, Jun. 2019, doi: 10.3390/app9112347.

[24] D.M Merkulov and I.V. Oseledets, “Empirical study of extreme overfitting points of neural networks,” vol. 64, no. 12, pp. 1527-1534, Dec. 2019, doi: 10.1134/s106422691912011.