Received: November 4, 2023. Revised: December 8, 2023.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Combination of BERT and Hybrid CNN-LSTM Models for Indonesia Dengue Tweets Classification

Wiwik Anggraeni[1]* Moch Farrel Arrizal Kusuma[1] Edwin Riksakomara[1]
Radityo P. Wibowo[1] Pujiadi[2] Surya Sumpeno[3]

[1]Department of Information Systems, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[2]Dengue Fever Eradication, Malang Regency Public Health Office, Malang, Indonesia
[3]Department of Computer Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
* Corresponding author's Email: wiwik@is.its.ac.id

Abstract: In the era of social media and online communication, the surveillance and classification of disease-related information in real-time is crucial. Twitter data on dengue-related are rarely used for classification, especially text-based classification in Indonesian. Even though, the classification of dengue-related news tweets is capable of being utilized for a variety of purposes. This study presents a novel approach to classifying dengue fever-related tweets in the Indonesian context, utilizing the potential of advanced language models and hybrid neural networks. The method proposed incorporates the advantages of two deep learning architectures: Bidirectional encoder representations from transformers indonesian-based (Indo-BERT) and a hybrid convolutional neural network-long short-term memory (CNN-LSTM) model which is still rarely used in this context. Indo-BERT, a pre-trained language model, extracts complex semantic and contextual information from text, enabling a deeper comprehension of dengue-related tweets. The Hybrid CNN-LSTM model processes textual data in tandem, extracting features via convolutional layers while maintaining temporal dependencies. To assure this model's effectiveness in the Indonesian context, a collection of dengue fever-related tweets in Indonesian was created and labeled for supervised learning. Experiments were conducted in multiple scenarios, and the results demonstrated that the combination of the Indo-BERT and Hybrid CNN-LSTM models had superior performance in classifying tweets about dengue illness into five labels, namely infected, awareness, informative, news, and others. The best models deliver an accuracy of 0.91, an F1-score of 0.90, a precision of 0.91, and a recall of 0.89. The hybrid model's efficacy surpasses that of previous approaches with an average difference in accuracy of 0.25, precision of 0.27, and recall of 0.26. It is anticipated that the Health Service can use the results of this classification to improve the dengue surveillance system with the initial data obtained from this classification.

Keywords: Social media, Twitter, Classification, BERT, Convolutional neural network, Long short-term memory.

1. Introduction

Dengue fever is caused by the dengue virus (DENV), which is transmitted to humans through the bite of an infected mosquito [1]. Dengue affects over half of the world's population, with an estimated 100-400 million cases occurring each year [2]. This disease is widespread globally in tropical and subtropical climates, primarily in urban and semi-urban regions [3]. Dengue fever has been a significant health concern in Indonesia for 47 years [2]. Indonesia is a tropical country. The number of dengue fever cases in Indonesia is significantly increasing, resulting in a number of deaths [4].

During the COVID-19 pandemic, Indonesians with dengue fever symptoms tended to avoid hospitals and other healthcare facilities out of fear. They typically use social media to research alternative treatment options. Consequently, there are numerous undocumented cases of dengue fever. The reality is that the number of instances registered in the

Health Service differs greatly from the actual number of cases in the field. This condition undoubtedly has an impact on the Health Service's inefficient distribution of resources and preventive actions.

Currently, social media is the preferred method of communication and news consumption. For event communication in real-time, social media has proved to be the most effective method. It is more real-time and immediate than traditional news sources [5]. Through social networks, millions of social media users express and communicate their thoughts [6]. Twitter has become an integral element of people's daily lives as a component of social media platforms [7, 8]. Twitter is a text-based social media platform that is excellent for investigating people's thoughts, opinions, and interests. This platform is reportedly a rich and effective source of information for text mining to obtain beneficial knowledge [9]. Twitter expedites the dissemination of information. The message can quickly reach a large audience when information is disseminated.

Previous research in a variety of disciplines has made extensive use of Twitter. Twitter has demonstrated its efficacy in the field of public health for supporting public health surveillance and identifying specific population groups [6, 10]. Tweet data is also processed to provide data sets that may be useful in the domains of epidemiology and infodemiology [11]. In the case of infectious diseases, Twitter data is commonly utilized to conduct sentiment analysis, detection, or prediction of disease transmission or trends on the topics of malaria [12], Zika [13], COVID-19 [14-17], and dengue fever [18-23].

There has been some research using Twitter data for dengue fever cases in recent years, but to the best of the author's knowledge, Twitter data on dengue fever cases are not widely used for classification purposes, particularly text-based classification in Indonesian. In fact, the results of the classification of dengue-related news tweets can be used for nowcasting functionality of current disease levels [20], forecasting [19] and early detection of an outbreak [21], as well as identifying the top users who are actively publishing relevant content about the topic [26, 27]. Twitter has been used in [22, 23] to track dengue fever cases in Indonesia. However, they made use of Twitter data by converting the location into time series data without processing the Twitter text.

There have been many approaches used to process Twitter data related to infectious diseases. Some existing approaches are Naive-Bayes [11] and decision trees [11, 24]. However, this approach is said to be less effective for unbalanced data [11] and sensitive to data outliers [24]. LSTM and CNN have also been used for the same purpose [25, 26]. However, LSTM is said to be less suitable for predicting text data because it is less able to predict words stored in long-term memory [26]. In addition, the performance of CNN models is influenced by their architectural settings [27] and it is often difficult to identify dominant information [28, 29].

Based on these conditions, this study aims to classify Twitter text data using a combination of the bidirectional encoder representations from transformers (BERT) and CNN-LSTM transfer learning models. Because this study utilized tweets written in Indonesian, the text classification was performed using the Indo-BERT model for Indonesian language representation. Using multiple classifier algorithm scenarios, the transfer learning procedure was carried out in order to identify the text classification composition that is most effective. CNN is good at predicting words in long-term memory, but not so good at capturing dominant information [28-30]. As a result, while LSTM is less capable of long-term prediction, it is effective in capturing dominant information [22, 28].

The main contributions of this study are as follows:

- Modeling Twitter text data associated with dengue illness in Indonesian with Indo-BERT
- Proposed a hybrid CNN-LSTM to classify Indo-BERT tweet data into informative, awareness, infected, news, and additional data groups.
- The tweet information used in Indonesia incorporates various language patterns, slang, and unique regional variations in social media communication.

This study not only contributes to the fields of natural language processing and disease monitoring, but it also demonstrates the potential of combining advanced language models with hybrid neural networks for classification and analysis of disease-related information on social media platforms in real time. Ultimately, it is anticipated that this classification of Twitter data will assist the Health Service in preparing valuable insights that can inform data-driven decision-making processes in order to limit the spread of dengue fever cases.

The remaining sections of this study are organized as follows. Section 2 provides a summary of the relevant prior research. In section 3, the dataset and methodology are described. In section 4, the results and discussion are discussed. The fifth section presented findings and recommendations for future research.

2. Related works

2.1 Utilization of Twitter data in the health sector

Twitter is a great text-based social media platform for investigating people's concerns and interest in diseases, particularly those transmitted by mosquitoes [9]. The transmission of disease is one of the many research areas [27] that can benefit from the content provided by social media platforms such as Twitter.

Previous research has classified health-related messages using Twitter data [20]. Twitter is said to be highly effective at detecting disease outbreaks from Twitter data. In addition, Twitter data is used to analyze public sentiment and its relationship to disease incidence, as well as to extract information regarding drug-disease relationships. Using Twitter data, [10] developed unsupervised models to detect spatial clusters that characterize high-risk regions. In addition, [11] has carried out the classification of tweet data in the Filipino language to determine whether or not a tweet is "infodemiological."

Regarding disease, [12] examined discourse by determining the polarity of categorized tweets and was able to produce recommendations for the treatment of malaria. Their study categorizes Twitter news in the English language into positive and negative categories. Twitter data was also used to investigate the contextual polarity of individuals' attitudes in social networks regarding malaria [31]. Twitter data has also been utilized by [13, 30] to investigate the short- and long-term effects of the Zika outbreak, in addition to malaria. Their study [32] conducted volumetric and co-occurring word analysis as well as examined trends in moderate Zika symptoms. In the COVID-19 disease, tweet data is utilized to present infodemic [14, 15], analyze public Covid sentiment [33], and detect or forecast COVID-19 trends [16].

Tweeter is said to be the most important corpus source for text mining in the case of dengue illness [9]. Tweet data has also been used to predict dengue cases at the local and national levels in Brazil [34] and in various locations of Indonesia [18, 19]. Tweet data is paired with case count data to forecast dengue fever cases during the coming time. The use of Twitter data in [19, 21] concentrates on the development of an intelligent indicator system from Twitter that uses sentiment analysis to detect dengue fever incidents in India. Additionally, [17] utilized polarity sentiment analysis to acquire data and extract pertinent information regarding dengue fever. The output is represented by the distribution of tweets per city and regions with the highest number of cases.

Previous studies have shown that Twitter may be used to collect information about disease spread. However, few of these studies have concentrated on categorizing news on Twitter in dengue fever cases. They are more concerned with categorizing emotions than with categorizing news. Similar to this, most processed Twitter texts use English or Filipino, but infrequently Indonesian. Twitter in Indonesian was processed by [18], but the output was predictions of the number of cases, not news classification.

This study is the first to classify Twitter data into five distinct categories. These include informative, awareness, infected, and news categories, among others. The text used on Twitter is in Indonesian. The results of this classification are derived from the prior classification, which categorized sentiment as positive or negative, neutral or not neutral, and others.

2.2 Approaches to Twitter text classification

Numerous approaches have been used to process text data pertaining to disease. Using the crimson hexagon ReadME algorithm, [12] determined the polarity of tweets and generated treatment recommendations for malaria. They use the red hexagon Opinion monitor approach to define the selected categories, but labeling in the training procedure is still performed manually [12, 30]. In addition, it has limitations in the number of tweets that can be retrieved and also needs to be linked to a trained algorithm to explore public opinion/sentiment in more depth [30]. [36] used Apache Hadoop and a sentiment analysis polarity approach to collect data and extract crucial information regarding dengue disease. This method collects data from social media using Apache Flume and extracts information using a hybrid polarity filtration algorithm in Apache Hive. Hadoop can handle problems with massive amounts of data, however it has limits for online analytical processing and decision support systems [36]. Hadoop is also said to require special expertise in management, configuration and maintenance needs, because it can be more complex.

Then, [32] investigated text mining using a word co-occurrence plot and hierarchical clustering algorithms. Their study only looked at the association between tweet themes and Zika sickness, and the results revealed a high match in terms of underlying theme identification. In terms of dengue fever, [24] classified tweet data into four categories: dengue and non-dengue, dengue combat, and dengue information. Their study made use of supervised algorithms including Naïve-Bayes, multinomial Naïve-Bayes, and J48 decision tree. Multinomial Naïve-Bayes had the greatest performance among the others, with an average accuracy of less than 72%, according to experimental results.

The Naïve-Bayes classification model has also been utilized by [11] to determine if tweets are "infodemiological" or not. Tweets written in the Filipino language. The model performance indicates that Naïve-Bayes can correctly classify 79.91% of tweets. This accuracy can be attributed to the limitations of Naïve-Bayes. In application, Naive-Bayes assumes that all features are independent of each other. This may not always be true in real world contexts. Additionally, it is less effective on imbalanced datasets [24]. In addition, classification of Twitter data was performed by [27] using the support vector machine approach, which produced the highest performance compared to other approaches. However, these SVMs require a considerable amount of training time and are less suitable for vast data sets [27]. SVM is also said to be sensitive to noise or outliers in the dataset.

As time progresses, researchers increasingly favor deep learning approaches that are deemed appropriate for large data volumes. Convolutional neural networks are used to classify text-based queries in [25], while recurrent memory networks (RMN) and long short-term memory (LSTM) are used for Twitter sentiment analysis in [26]. According to [25], however, CNNs require configuration settings to enhance the accuracy of text-based classification. Furthermore, LSTM is said to be deficient in terms of perplexity on three large datasets (Italian, German, and English) [26] and is less suited for predicting text data because it is less capable of predicting words held in long-term memory. In addition, [27] demonstrates that CNN settings can affect model performance. Optimal parameters are determined based on the language. Testing on a variety of data verifies the optimization and confirms the transferability of the optimal parameters [25].

Then, [35] used LF, CNN, LSTM, and Bi-LSTM, each of which was combined with LF, to classify Twitter into three sets: multi-class, neutral vs. non-neutral, and positive vs. negative messages. Experimental results indicate that the classification (for multi-class classification) of infectious diseases, such as dengue, is still performing below 60% on average. The results of the binary classification of neutral and non-neutral tweets reveal an average accuracy of 57.81% across all methods used. For positive and negative tweets, the average accuracy is 67.89%; and the CNN algorithm has the maximum accuracy with an average accuracy of 70.52%. This less-than-optimal performance can be caused by LSTM which can be susceptible to overfitting, especially if it is not tuned properly or if the training data is limited. Although LSTMs are designed to handle long-range dependencies, it may still be difficult to understand very long relationships in sequences [25]. Similarly, CNNs are less effective at modeling long-term context in sequential data, such as in natural language processing tasks [26].

For sentiment and complaint identification, [37] uses word embeddings supplemented with deep learning models such as DNN, LSTM, Bi-LSTM, and CNN. The BERT model was incorporated to the classification system in this study. The results demonstrate that the BERT model outperformed other integrated deep learning benchmark word embeddings models (DNN, LSTM, Bi-LSTM and CNN) [37].

The BERT model is also being used in a variety of studies [37]. This could be a follow-up to [12], who suggested that future text data processing studies should employ NLP models. Model creation is becoming more popular as a means of improving classification performance. According to [27], BERT has an excellent performance for various NLP applications. Their study used BERT to do sentiment analysis on tweets in both English and Italian, and the results were substantial. BERT is stated to be accessible and successful in a number of languages, including English [27], Italian [27], Spanish [38], and Indonesian [28].

This study proposes combining CNN and LSTM, which are applied to the Indonesian language-based BERT model, or what is commonly known as IndoBERT, based on the approach's advantages and disadvantages. Even though BERT has been utilized extensively, it can have varying levels of difficulty and efficacy for each language [27].

3. Materials and method

3.1 Dataset

This study utilized Twitter data in the Indonesian language for the time period spanning 1 January 2013 to 31 August 2023. This study analyzed Twitter data in Indonesian using the most popular keywords related to dengue fever that were frequently used by Indonesians. This data set was comprised of a diversity of language patterns, slang, and regionally specific variations in social media communication in Indonesia.

Received: November 4, 2023. Revised: December 8, 2023.

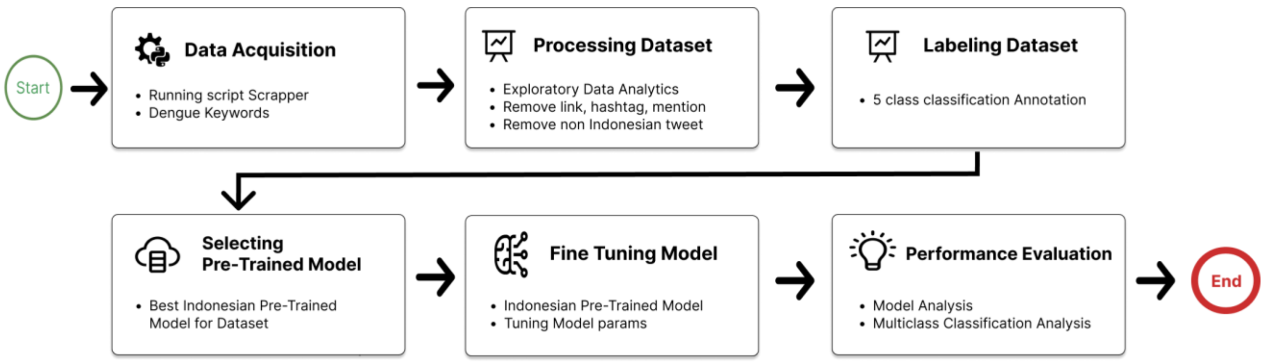

Figure. 1 The framework of this study

3.2 Method

This study divided previously collected Twitter data into five categories. For classification, an LSTM-CNN hybrid model was used. The performance of the model was compared to that of several existing approaches, including Naïve-Bayes [11, 21], SVM [27], Decision Tree [24], LSTM [35] [26], and CNN [28, 33, 34]. Fig. 1 depicts the steps that were taken to accomplish the objectives of this study.

3.2.1. Data acquisition

The collection of data from social media, specifically Twitter in Indonesia, was a primary focus of this study. We used a variety of approaches, such as the Twitter API and web scraping, to retrieve data. As tweet metadata, the Twitter API allowed for the retrieval of a variety of information. This approach had a limitation in that the most recent data could only be collected for up to seven days. In order to circumvent this issue, selenium was employed, which was one method for retrieving various data from a time range specified by keywords.

The Indonesian keywords used in this study are *demam berdarah* (or dengue fever in English), dengue, *gejala DBD* (or dengue fever symptoms in English), *gejala demam berdarah* (or dengue fever symptoms in English), *nyamuk DBD* (or dengue mosquito in English), *sakit DBD* (or dengue fever in English), *ciri-ciri demam berdarah* (or the characteristics of dengue fever in English), *pengertian demam berdarah* (or the definition of dengue fever in English), and *poster demam berdarah* (or dengue fever poster in English). The scraping process was devised to enable account authentication, keyword-based searches, and scrolling to obtain massive amounts of user tweets. The extraction results were in CSV format and contained a total of 6899 raw data that were prepared and ready to join the model training phase.

3.2.2. Processing dataset

This stage of preprocessing aimed to enhance data quality. In this study, several processes were carried out, such as converting all terms to lowercase, removing duplicate metadata, and removing all mentions, hashtags, and links that had nothing to do with classification, as well as removing all advertisements that were placed on Twitter intentionally. This was possible due to the fact that advertisements posted by Twitter differed slightly from tweet data created by humans, for which metadata such as posting date and ID do not exist. Therefore, this made it simpler to categorize advertisements created by Twitter through tweet posts.

The process of retrieving data from different keyword topics made it possible for duplicate tweets to reappear in other search processes with different keywords. Consequently, this had to be resolved by modifying the username and tweet content. The procedure of removing duplicate tweets was conducted by matching the tweet's ID, user name, and tweet. The duplicate was eliminated if the matching results were identical.

Indonesian tweets were then filtered as the next step in this stage. This was used to filter Malay comments with a high degree of grammatical and lexical similarity. Given that the pre-trained model that would be used during training was trained using an Indonesian corpus, this tweet filtering procedure was required to obtain tweets that were composed entirely of Indonesian.

The filtering was done with Google translate, and the results were checked again when the data was labeled manually. As a result, if there was still Malay language coming in, the annotator would erase it. Finally, from 6899 raw data after cleaning, 2790 data were obtained that were ready to be used for model training. Fig. 2 shows the processes that follow the results of data cleaning.

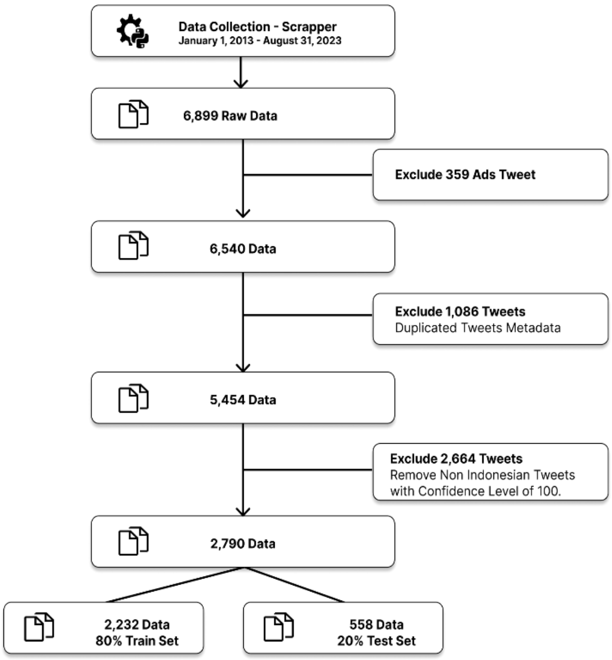

Figure. 2 The Stage of data cleansing

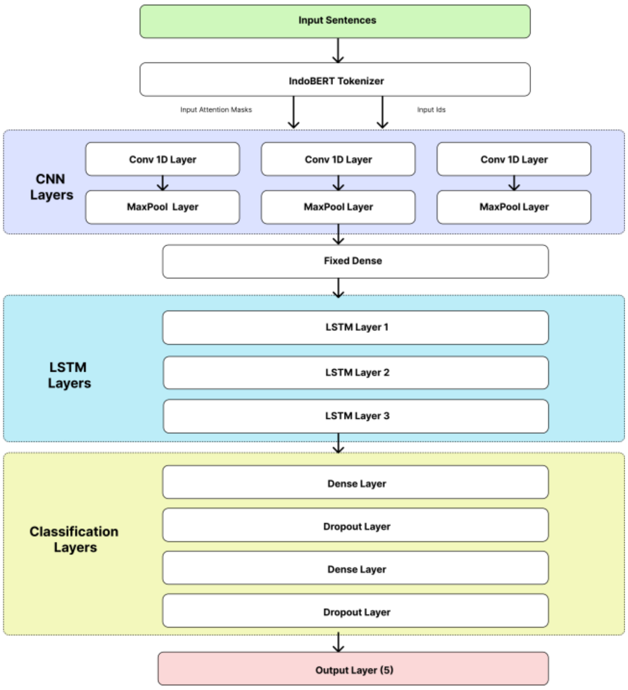

Figure. 3 Hybrid model architecture

3.2.3. Data annotation/labeling

Five classes were then manually assigned to the cleaned tweet data. The programs utilized for this study were Infected, Awareness, Informative, News, and Others. In order to establish the credibility of the annotator, the labeling procedure was carried out by native Indonesian speakers with formal education. The model was then divided into two sections, the train set and the test set, with a total of 2232 and 558 data, respectively.

3.2.4. Selecting pre-trained model

BERT is a language representation model that was initially developed by researchers at Google AI Language. BERT is designed to train the bidirectional representation of unlabeled text in depth by considering both left and right contexts for each text layer [27]. BERT uses a masked language model (MLM) to improvise. The MLM is trained to randomly mask some word tokens from its input and attempts to predict the actual vocabulary token ids of the masked words based on the sentence context. BERT depends on transformers that can learn contextual relationships between words in a text via an attention mechanism [35, 36].

When a sentence is processed using a BERT model, many factors occur to aid the model's word prediction. There are three encoding structures, which are Token embeddings, Segment Embeddings, and Position Embeddings. Token embeddings entail supplying distinguishing identities to differentiate words and context, such as CLS as a preposition and sep as a connecting pronoun between words. Segment embeddings provide specific identifiers for distinguishing context within sentences. Position embedding makes it possible to determine the position of words within a sentence.

Because the tweets used in this study are in Indonesian, the Indo-BERT algorithm was chosen based on [28]. The Indo-BERT model is the Indonesian adaptation of the BERT model. It trains the model with over 220M words collected from three primary sources, including Wikipedia in Indonesia (74 million words), Kompas, Tempo, and Liputan6 news stories (55M words in total) (90 million words), Indonesian web corpus. The model was trained for 2.4 million steps (180 epochs) with a final perplexity of 3.97 (similar to the English-based BERT).

3.2.5. Fine tuning classification model

To complete the classification assignment into five categories, this study employed three approaches. The first and second approaches were LSTM and CNN-based, respectively. Meanwhile, the second strategy was a hybrid approach that utilized a combination of CNN and LSTM. The utilized model architecture is depicted in Fig. 3. The used language models were IndoBERT-base-p1 and IndoBERT-large-p2, both of which were trained using the Indonesian language corpus. In scenario 1, the pre-trained base model was utilized, while in the remaining scenarios, the pre-trained large model was utilized. In this study, a number of model scenarios were evaluated.

In general, the steps in LSTM can be explained mathematically as follows:

The first part of the mechanism is to enable LSTM model to forget the irrelevant information. This is achieved by means of forget gate f_t and is given as Eq. (1) where x_t here is the current input while h_{t-1} is the previous timesteps' output after passing through an output gate.

$$f_t = \sigma(W_f \, . \, [x_t, h_{t-1}] + b_f) \quad (1)$$

The second portion, determining significant information to be stored in the current cell state, is handled by input gate i_t. This is done by Eq. (2). W_i is input weight for the update gate, U_i is recurrent weight for the update gate, b_i shows the bias weight for the update gate, and σ shows the activation function.

$$i_t = \sigma(W_i . [h_{t-1}, x_t] + b_i) \quad (2)$$

$$S_t = tanh(W_S \, . \, [h_{t-1}, x_t] + b_S) \quad (3)$$

The combination of Eq. (2) and Eq. (3) with old cell state determines the new cell state Eq. (4).

$$S_t = f_t \odot S_{t-1} \oplus i_t \tilde{\odot} S_t \quad (4)$$

The final step to determine the output (o_t) of a memory cell is given as Eq. (5) where $\odot$ and $\oplus$ denote element-wise multiplication and element-wise summation operation, W_i, W_f, W_S that represent weights, b_i, b_f, b_S represent bias values, x_t represents the input at timestamp t, σ state the activation function and h_t represents the hidden state at timestamp t.

$$o_t = \sigma(W_o . [h_{t-1}, x_t] + b_0) \quad (5)$$

$$h_t = o_t \odot tanh(c_t) \quad (6)$$

In addition, the CNN model can be explained mathematically as follows:

Suppose that we have some square neuron layers of size 'N×N' which are followed by convolutional layer. If we use filter ω of size 'm×m', our convolutional layer output will be of size '(N−m+1)×(N−m+1)'. In order to compute the pre-nonlinearity input to some unit x_{ij}^l in our layer, we need to sum up the weighted by the filter components from the previous layer cells as Eq. (7). Then, the convolutional layer applies its nonlinearity as Eq. (8).

$$x_{ij}^l = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} \omega_{ab} Y_{i+a,j+b}^{l-1} \quad (7)$$

$$Y_{ij}^l = \sigma(x_{ij}^l) \quad (8)$$

Meanwhile, for backward propagation, first determine the gradient value for each weight using the chain rule. All arguments in which the variable appears must contribute to the chain rule as shown in Eq. (9). For determining the gradient, the values of $E(x_{ij}^l)$ need to be identified first. The gradients are quite simple to calculate by applying the chain rule as shown in Eq. (10) and Eq. (11). Suppose E is error function, N shows the size of neuron layer, m is size of filter ω, $\frac{\partial E}{\partial y_{ij}^l}$ is the error at the current layer and σ'(x) is the derivative of the activation function.

$$\frac{\partial E}{\partial \omega_{ab}} = \sum_{i=0}^{N-m} \sum_{j=0}^{N-m} \frac{\partial E}{\partial x_{ij}^l} \frac{\partial x_{ij}^l}{\partial \omega_{ab}} = \sum_{i=0}^{N-m} \sum_{j=0}^{N-m} \frac{\partial E}{\partial \omega_{ij}^l} Y_{(i+a)(j+b)}^{l-1} \quad (9)$$

$$\frac{\partial E}{\partial x_{ij}^l} = \sum_{i=0}^{N-m} \sum_{j=0}^{N-m} \frac{\partial E}{\partial y_{ij}^l} \frac{\partial y_{ij}^l}{\partial x_{ij}^l} = \sum_{i=0}^{N-m} \sum_{j=0}^{N-m} \frac{\partial E}{\partial y_{ij}^l} \frac{\partial}{\partial x_{ij}^l} (\sigma(x_{ij}^l)) \quad (10)$$

$$\frac{\partial E}{\partial y_{ij}^{l-1}} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} \frac{\partial E}{\partial x_{(i-a)(j-b)}^l} \frac{\partial x_{(i-a)(j-b)}^l}{\partial y_{ij}^{l-1}} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} \frac{\partial E}{\partial x_{(i-a)(j-b)}^l} \omega_{ab} \quad (11)$$

This study implemented an architectural model using the TensorFlow Keras library and DGX A 100 graphics computing to complete the training procedure for the model. This model was designed to perform classification tasks involving five classes. Given that this is a multi-class classification, "sparse categorical cross entropy" was utilized as one of the loss metrics with a learning rate of 1e-5 in this study. 20 epochs were designated as the number of training iterations. The number of epochs was determined based on observations of losses that had begun to become stationary, resulting in minimal class-wide model learning.

3.2.6. Performance evaluation

Accuracy, F1-score, precision, and multiclass support were utilized to evaluate the performance of the Indo-BERT model in various experimental scenarios [29, 40], as illustrated by Eq. (12) to Eq. (15). Assumptions A, B, C, D are class labels, total number represents a lot of data, True A represents the

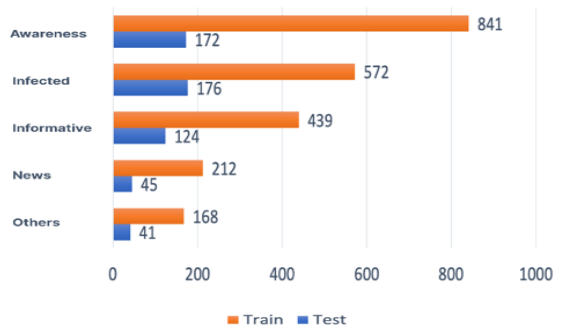

Figure. 4 Polarization of tweet labeling results

actual amount of data labeled class A and correctly predicted as class A labels.

$$Accuracy = \frac{number\ of\ true\ classifications}{total\ number\ of\ data} = \frac{True\ A + True\ B + True\ C + True\ D}{total\ number} \quad (12)$$

$$Precision(A) = \frac{\Sigma x(c = A|\hat{c} = A)}{\Sigma x_{(\hat{c}=A)}} = \frac{True\ A}{\Sigma\ Pred\ c{=}A} \quad (13)$$

$$Recall(A) = \frac{\Sigma x(\hat{c} = A|c = A)}{\Sigma x_{(c=A)}} = \frac{True\ A}{\Sigma\ Actual\ c{=}A} \quad (14)$$

$$F1 - Score = \frac{2 \text{ x } (Recall \text{ x } Precision)}{(Recall + Precision)} \quad (15)$$

4. Result and discussion

The primary goal of this study is to classify tweets about dengue fever in Indonesian so that it can be determined whether the tweet indicates that someone is infected with dengue fever or whether it is news, informative, awareness-raising, or other categories.

Fig. 4 interprets the results of label polarization. The results of this study's data polarization indicate that the Awareness label dominates all other labels with a total of 1017 tweets, or 36.31 percent. The percentage of tweets that were infected was 26.81%, followed by the percentage of tweets that were informative, which was 20.18 %. Then, news-only ----tweets account for up to 9.21%, followed by all other tweets, which account for up to 7.49%. This composition also demonstrates that the data used to create this classification are unbalanced data. In spite of this, it is anticipated that the proposed hybrid model will have comparable efficacy to previously utilized models. In the meantime, the evaluation samples and target labels are listed in Table 1. Then, the tweets were grouped using a hybrid CNN and LSTM by applying the seven scenarios enumerated in Table 2.

Table 3 shows that the best scenario of the proposed classification model is densest in scenario 7 with an architecture consisting of the first CNN layers with 512 filters and kernel_size 3, MaxPooling1D with pool size 4), second CNN layers with 256 filters and kernel_size 5, MaxPooling1D with pool size 4, and the third CNN layer with 128 filters and kernel_size 7, MaxPooling1D with pool size 3. Meanwhile, the first LSTM layer has an input of 128 units, a Dense Layer of 64 units, and a dropout of 0.2. The second layer of LSTM is comprised of a Dense Layer with 32 units and a deviation of 0.2. Consequently, all ensuing operations are carried out in scenario 7.

The model from Scenario 7 is then utilized to classify testing data, and its efficacy is displayed in Table 4. Table 4 is a confusion matrix for each label for which the classification performance of the model is, on average, quite excellent.

The average accuracy of the model's predictions for each label is 91%. At 96%, informative labels were the most competent, followed by suspicious and news labels at 93%. The Others label displays the least significant value. The less significant value condition is caused by the fact that the used tweet data is unbalanced and the number of available "others" data classifications is relatively low. This is consistent with what [38, 39] stated, namely that classification of unbalanced data presents a challenge in classification tasks, particularly when it comes to achieving high performance [39]. In addition, this condition is consistent with [39, 41], which state that certain models are extremely sensitive to this circumstance of unbalanced data. Unbalanced data will also result in a classification with inferior performance compared to balanced data.

The proposed model has, in general, a high sensitivity for comprehending each class. Sensitivity was 89% on average. 97% of infected labels were successful, followed by 96% of news articles. The value for awareness was 91%, the value for informative was 88%, and the value for the others label was 73%. This condition is also supported by the precision and F1-score values, which demonstrate a similar correlation between each label and each value. This demonstrates that the proposed hybrid model is capable of capturing syllables stored in long-term memory, which is an advantage of the CNN-LSTM model. CNN is effective at predicting words retained in long-term memory, but it is less adept at capturing dominant information. Therefore, LSTM is less able to predict the long term, but it is effective at capturing dominant information. This corroborates what [22] and [28] said about LSTM as well as the opinions of [28-30] about CNN.

Table 1. Examples of reviews and generated labels

Tweets (Indonesian)	Tweets (English)	Label
Gw tuh kemaren kena dbd trs trs kan disuruh rawat inep y tp gw gamau krn takut diinfus sakit jd 4 hari kmrn tetep di rumah tp ambil darah konsul gt alhamdulillah kt dokternya udah gaperlu ke rs lagii yey	I was diagnosed with dengue fever yesterday, and I was initially advised to be hospitalized, but I didn't want to because I was afraid of the pain from the IV. So, for the past four days, I stayed at home and had blood tests and consultations. Alhamdulillah, the doctor said I no longer need to go to the hospital. Yey!	Infected
who menyebut tahun 2023 dan 2024 akan ditandai dengan fenomena el nino fenomena tersebut dikatakan dapat meningkatkan penularan virus seperti demam berdarah dan arbovirus	WHO has been mentioned that the years 2023 and 2024 will be marked by the El Nino phenomenon, which is said to potentially increase the transmission of viruses such as dengue fever and arboviruses.	News
pagi teman teman semua gimana puasanya masih lancar btw di tempat kalian masih sering hujan nggak kalau di tempatku sih masih alhamdullilah bikin adem tapi waspada juga banyak air menggenang diselokan dan bikin banyak nyamuk nih	Good morning, everyone! How's your fasting going? By the way, do you still experience frequent rainfall in your area? In my location, it's still happening, and thankfully, it keeps the weather cool. However, we need to be cautious as there's a lot of standing water in the streets, which can lead to a high mosquito population.	Awareness

Table 2. Experiment scenario

Scenario	Characteristics of Hybrid Model
Scenario-1	input LSTM (32 units), Bidirectional, GlobalMaxPool1D, dense layers (32 units), output dense layers, base pretrained.
Scenario-2	input LSTM (32 units), Bidirectional, GlobalMaxPool1D, dense layers (32 units), output dense layers, large pretrained
Scenario-3	input LSTM (128 units), Dense layer (64 units), Dropout (0.2), Dense layer (32 units), output dense layers, large pretrained
Scenario-4	CNN layer (filters=128; kernel size=5), LSTM (128 units), Dense Layer (64 units), Dropout (0.2), Dense Layer (32 units), Dropout (0.2), Output
Scenario-5	CNN layers (filters=256; kernel_size=3), MaxPooling1D (pool size=4), CNN Layers (filters=128; kernel_size 3), GlobalMax_ Pooling1D (pool_size=3), Reshape layers (128, 1), LSTM Input (128 units), Dense Layer (64 units), Dropout (0.2), Dense Layer (32 units), Dropout(0.2), Output
Scenario-6	CNN layers (filters=128; kernel_size=3), MaxPooling1D (pool size=4), CNN Layers (filters=128; kernel_size=4), MaxPooling1D (pool_size=3), CNN Layers (filters=128; kernel_size=5),MaxPoo_ ling1D (pool_size=3), LSTM Input (128 units), Dense Layer (64 units), Dropout (0.2), Dense Layer (32 units), Dropout (0.2), Output
Scenario-7	CNN layers (filters=512; kernel_size=3), MaxPooling1D (pool size=4), CNN Layers (filters=256; kernel_size=5), MaxPooling_ 1D (pool_size=4) , CNN Layers (filters=128; kernel_size=7), MaxPooling1D (pool_size=3), LSTM Input (128 units), Dense Layer (64 units), Dropout (0.2), Dense Layer (32 units), Dropout(0.2), Output.

Table 3. Performance of the proposed training model

Scenario	Accuracy	F1-Score	Precision	Recall
Scenario 1	0.80	0.76	0.77	0.75
Scenario 2	0.83	0.82	0.86	0.79
Scenario 3	0.81	0.80	0.82	0.80
Scenario 4	0.83	0.80	0.80	0.81
Scenario 5	0.83	0.79	0.88	0.76
Scenario 6	0.86	0.84	0.83	0.82
Scenario 7	0.91	0.90	0.91	0.89

Table 4. Confusion matrix of scenario 7

Labels	Precision	Recall	F1-Score	Support
Awareness	0.87	0.91	0.89	172.00
Infected	0.93	0.97	0.95	176.00
Informative	0.96	0.88	0.92	124.00
News	0.93	0.96	0.95	45.00
Others	0.86	0.73	0.79	41.00

Table 5. Classification report of scenario 7

Labels		Predictive				
		Awareness	Infected	Informative	News	Others
Actual	Awareness	**156**	9	3	2	2
	Infected	5	**171**	0	0	0
	Informative	11	0	**109**	1	3
	News	2	0	0	**43**	0
	Others	6	4	1	0	**30**

The classification report, shown in Table 5, also supports this high model performance. The numbers on the main diagonal in Table 5 reveal that the expected and actual results are dominated by the main diagonal. This demonstrates that the model was successful in predicting. Therefore, even if the data conditions for each group of labels are uneven, the model can still generate decent results.

Furthermore, it has been discovered that the model still has errors in class prediction. Table 6 shows a few examples of inaccurate classification results. In the first and second columns of Table 6, several instances of tweet data that were incorrectly predicted by the model are displayed. The first column contains the original statement in Indonesian, while the second column provides an English translation. Most of these erroneous predictions are the result of statements that are not explicit or are also used as parables when there is a dearth of comparable data in the dataset. This causes the model to continue to struggle with classifying specific conditions.

For the purpose of determining whether the proposed model improves classification performance, the model is contrasted to several other approaches that have been utilized in the past. The employed comparison method is as described in subchapter 3.2. The dataset used in this performance comparison is the same dataset. The dataset used as input from all methods is twitter data in Indonesian for the period January 1, 2013 to August 31, 2023 as described in subchapter 3.1. This data is obtained from the data acquisition process mentioned in subchapter 3.2.1.

The comparative performance results are shown in Table 7. S1 to S7 in Table 7 represent the experimental scenarios 1 through 7 described in Table 2. The var_smoothing parameters used in the Naïve-Bayes experiment are 1e-09. While the parameters for the decision tree are criterion='gini', splitter='best', max_depth=None, min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features=None, random_state=123, max_leaf_nodes=None,min_impurity_decrease=0.0, class_weight=None, ccp_alpha=0.0.

For the SVM, loss='hinge', penalty='l2', alpha=0.0001, l1_ratio= 0.15, fit_ intercept=True, max_iter=1000, tol=0.001, shuffle =True, verbose=0, epsilon=0.1, n_jobs= None, learning_rate='optimal', eta0=0.0, power_t =0.5, early _stopping=False, validation_ fraction =0.1, n_iter_ no_change=5, class_weight=None,warm_start=False, average=False. While the parameters used for others are shown in Table 2.

Table 6. Sample of Miss-classification

Tweet		Label	
Indonesian	English	Prediction	Actual
Bagi sobat tropmed yang ingin bertanya silakan tulis di kolom komentar yaa	For fellow tropical medicine enthusiasts who have questions, please feel free to write them in the comment section.	Awareness	Others
Selamat berbahagia kevin dan pewaris mnc group perlu dikit drama ajah untuk mengelabui publik gak ikut all england alasannya lagi sakit dbd eh taunya mau married si kevin	Congratulations, Kevin, the heir of MNC Group. A little drama is needed to deceive the public - not participating in the All England because of dengue fever, but it turns out he's getting married.	Awareness	Others
Aku di ilmu ekonomi kmrn dpt statistika 2 smt aja serasa sakit demam berdarah tipes ini lagi kamu mau ambil statistika bareng teksip lagi jgnnn.	I was in economics last semester, and getting Statistics 2 felt like having dengue fever. I don't want to take statistics again with you this semester. Please, no.	Infected	Others
Setau saya lapor ke pihak desa puskesmas kalo ada kejadian demam berdarah soalnya nanti dari puskesmas yang melakukan fogging kalo jasa layanan swasta gitu kurang tau yaa	As far as I know, it's important to report cases of dengue fever to the village or sub-district health center (Puskesmas) because they are responsible for conducting fogging and related services. Private service providers may not be well-versed in handling such cases.	Awareness	Informative

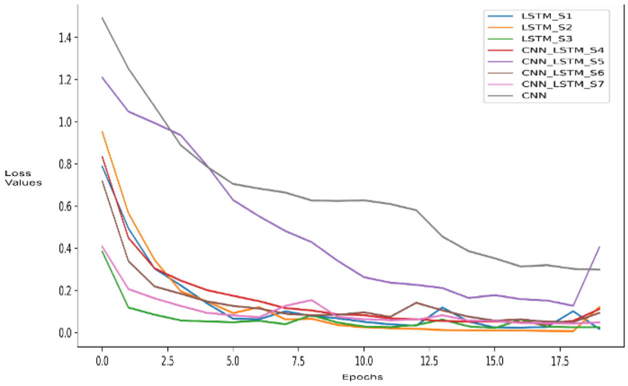

Figure. 5 Loss value movement of every epoch

Table 7. Performance comparison of the hybrid model

Method	Accuracy	F1-Score	Precision	Recall
Naïve Bayes [11][24]	0.28	0.23	0.3	0.28
Support Vector Machine [27]	0.28	0.20	0.23	0.28
Decision Tree [24]	0.30	0.29	0.29	0.30
LSTM - S1 [20][35]	0.80	0.76	0.77	0.75
LSTM - S2 [20][35]	0.83	0.82	0.86	0.79
LSTM - S3 [20][35]	0.81	0.80	0.82	0.80
CNN [27][35]	0.80	0.64	0.61	0.69
CNN-LSTM - S4	0.83	0.80	0.80	0.81
CNN-LSTM - S5	0.83	0.79	0.88	0.76
CNN-LSTM - S6	0.86	0.84	0.83	0.82
CNN-LSTM - S7	**0.91**	**0.90**	**0.91**	**0.89**

The average performance of the proposed hybrid model is greater than that of its predecessor, as shown in Table 7. CNN-LSTM performs the best in scenario 7 with a 91% accuracy rating. Meanwhile, scenario 2's LSTM achieves an accuracy of 82%. The same holds true for the precision, recall, and F1 scores. The hybrid model proposed has the best efficacy in comparison to other models. The accuracy of CNN is 80%, 11% less than the proposed model. This is consistent with what [28] states, namely that CNN can be used for text-based classification, but the optimal parameter configuration must be determined. This is also consistent with what [30] states. According to [29], CNN can achieve an average accuracy of 70.52 % for binary classification with the appropriate configuration parameters.

Then, for LTSM, in the three cases presented, LSTM has an average accuracy of 81%, which is still lower than the hybrid model's average performance. This is due to the length of the tweet text, which demands a long memory to remember. This is consistent with the findings of [22, 28], who found that LSTM is less capable of predicting words retained in long-term memory. In addition, LSTM frequently exhibits overfitting, as evidenced by the Loss Graph in Fig. 5.

Naïve-Bayes has a 28% accuracy rate. This is due to the fact that the classification in this instance is a multiclass classification (more than two classes) and the data is unbalanced. This is corroborated by [13, 23] as well. Similarly, SVM and Decision Tree have the same accuracy as Naïve-Bayes. This low SVM efficacy may be a result of the extensive amount of data employed. This is consistent with what [30] states, namely that SVM is less suited for large data sets. According to [24], the performance of decision trees is inferior to that of Naïve-Bayes for tweeter classification.

In addition to performance comparisons in the form of metrics, Fig. 5 also depicts a comparison of the models in Table 7 based on loss value reduction. Fig. 5 shows that the hybrid CNN_LSTM model converges faster than the other models. Even though the hybrid model has a greater loss value when it starts off, it eventually finds a stable situation faster than CNN. CNN continues to fluctuate as the epoch grows.

5. Conclusion and future works

The experimental results demonstrate that the hybrid CNN-LSTM approach provides superior performance under all conditions for the hybrid CNN-LSTM approach model. With average values of 89% and 93% for sensitivity and precision, respectively, the model is able to classify classes effectively. The application of data cleansing and training with a larger corpus of data that has been pre-trained yields impressive results in support of improved model classification.

Pretrained Indo-BERT performs well when used to classify text. Based on the fact that Indo-BERT was trained specifically for Indonesian, it has a deeper understanding, a richer vector representation, and has undergone a succession of intensive training sessions in the past. The quantity and quality of the training data is one of the most important determinants of the model's ability to perform its assigned tasks. In order for Indo-BERT to continue training in classifying texts into five classes, this can be accomplished correctly by implementing proper data cleaning.

Experiments have revealed that there are still errors, particularly in labels with limited data. This classification is based on unbalanced information. In the subsequent study, the combination of Indo-BERT and hybrid CNN_LSTM will therefore be applied to the data after a balancing procedure. It is anticipated that this will enhance the performance of classification.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

Wiwik Anggraeni: conceptualization, methodology, formal analysis, writing—original draft preparation and editing. Moch Farrel Affrizal Kusuma: data curation, experiments, formal analysis. Edwin Riksakomara: conceptualization, validation, writing—review. Radityo P. Wibowo: validation, formal analysis, writing—review, Surya Sumpeno: supervision, conceptualization, writing—review. All authors read and approved the final manuscript.

Acknowledgments

We would like to express our gratitude to the Ministry of Research, Technology, and Higher Education of the Republic of Indonesia for providing research funding with contract No. 112/E5/PG.02.00.PL/2023. Our gratitude also goes to Malang Regency Public Health Services for their assistance.

References

[1] "WHO | What is dengue?", WHO. Accessed: Apr. 19, 2020. [Online]. Available: http://www.who.int/denguecontrol/disease/en/

[2] "Dengue and severe dengue", Accessed: Oct. 16, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue

[3] WHO, "WHO | Dengue guidelines for diagnosis, treatment, prevention and control: new edition", WHO. Accessed: Mar. 28, 2019. [Online]. Available: https://www.who.int/rpc/guidelines/9789241547871/en/

[4] B. H. Ditjen. P2P, "Kesiapsiagaan Menghadapi Peningkatan Kejadian Demam Berdarah Dengue Tahun 2019 |Preparedness for Facing the Increased Incidence of Dengue Hemorrhagic Fever in 2019| Direktorat Jendral P2P", Accessed: Sep. 20, 2023. [Online]. Available: http://p2p.kemkes.go.id/kesiapsiagaan-menghadapi-peningkatan-kejadian-demam-berdarah-dengue-tahun-2019/

[5] R. Prasad, A. U. Udeme, S. Misra, and H. Bisallah, "Identification and classification of transportation disaster tweets using improved bidirectional encoder representations from transformers", *International Journal of Information Management Data Insights*, Vol. 3, No. 1, p. 100154. 2023.

[6] Y. Chen and Z. Zhang, "An easy numeric data augmentation method for early-stage COVID-19 tweets exploration of participatory dynamics of public attention and news coverage", *Information Processing and Management*, Vol. 59, No. 6, p. 103073, 2022.

[7] A. Sharma, K. Sanghvi, and P. Churi, "The impact of Instagram on young Adult's social comparison, colourism and mental health: Indian perspective", *International Journal of Information Management Data Insights*, Vol. 2, No. 1, p. 100057, 2022.

[8] M. Mahdikhani, "Predicting the popularity of tweets by analyzing public opinion and emotions in different stages of Covid-19 pandemic", *International Journal of Information Management Data Insights*, Vol. 2, No. 1, p. 100053, 2022.

[9] S. Q. Ong, M. B. M. Pauzi, and K. H. Gan, "Text mining in mosquito-borne disease: A systematic review", *Acta Troprica*, Vol. 231, p. 106447, 2022.

[10] R. C. S. N. P. Souza, R. M. Assunção, D. M. Oliveira, D. B. Neill, and W. Meira, "Where did I get dengue? Detecting spatial clusters of infection risk with social network data", *Spatio-Temporal Epidemiology*, Vol. 29, pp. 163–175, 201.

[11] K. Espina, M. R. J. Estuar, D. J. S. IX, R. J. E. Lara, and V. C. D. L. Reyes, "Towards an Infodemiological Algorithm for Classification of Filipino Health Tweets", *Procedia Computer Science*, Vol. 100, pp. 686–692, 2016.

[12] J. Boit and O. E. Gayar, "Topical Mining of Malaria Using Social Media. A Text Mining Approach", *Faculty Research and Publication*, 2020, [Online]. Available: https://scholar.dsu.edu/bispapers/222

[13] L. Safarnejad, Q. Xu, Y. Ge, A. Bagavathi, S. Krishnan, and S. Chen, "Identifying Influential Factors in the Discussion Dynamics of Emerging Health Issues on Social Media: Computational Study", *JMIR Public Health and Surveillance*, Vol. 6, p. e17175, 2020.

[14] R. Kumari, N. Ashok, T. Ghosal, and A. Ekbal, "Misinformation detection using multitask learning with mutual learning for novelty detection and emotion recognition", *Information Processing & Management*, Vol. 58, No. 5, p. 102631, 2021.

[15] J. Agley and Y. Xiao, "Misinformation about COVID-19: evidence for differential latent

profiles and a strong association with trust in science", *BMC Public Health*, Vol. 21, No. 1, p. 89, 2021.

[16] W. Huang, B. Cao, G. Yang, N. Luo, and N. Chao, "Turn to the Internet First? Using Online Medical Behavioral Data to Forecast COVID-19 Epidemic Trend", *Information Processing & Management*, Vol. 58, No. 3, p. 102486, 2021.

[17] N. B, A. Ghani, S. Hamid, M. Ahmad, Y. Saadi, N. Z. Jhanjhi, M. A. Alzain, and M. Masud, "Tracking Dengue on Twitter Using Hybrid Filtration-Polarity and Apache Flume", *Computer System Science & Engineering*, Vol. 40, No. 3, pp. 913–926, 2021.

[18] A. L. Ramadona, Y. Tozan, L. Lazuardi, and J. Rocklöv, "A combination of incidence data and mobility proxies from social media predicts the intra-urban spread of dengue in Yogyakarta, Indonesia", *PLoS Neglected Tropical Disease*, Vol. 13, No. 4, p. e0007298, 2019.

[19] W. Anggraeni, E. M. Yuniarno, R. F. Rachmadi, Pujiadi, and M. H. Purnomo, "A Sparse Representation of Social Media, Internet Query, and Surveillance Data to Forecast Dengue Case Number using Hybrid Decomposition-Bidirectional LSTM", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 5, pp. 209–225, 2021, doi: 10.22266/ijies2021.1031.20.

[20] O. Șerban, N. Thapen, B. Maginnis, C. Hankin, and V. Foot, "Real-time processing of social media with SENTINEL: A syndromic surveillance system incorporating deep learning for health classification", *Information Processing & Management*, Vol. 56, No. 3, pp. 1166–1184, 2019.

[21] V. Lampos and N. Cristianini, "Tracking the flu pandemic by monitoring the social web", In: *Proc. of 2nd International Workshop on Cognitive Information Processing*, pp. 411–416, 2010.

[22] W. Anggraeni, E. M. Yuniarno, R. F. Rachmadi, and M. H. Purnomo, "Fuzzy C-Means and Social Network Analysis Combination for Better Understanding the Patient-based Spread of Dengue Fever with Climate and Geographic Factors", *International Journal of Intelligent Engineering and Systems*, Vol. 15, No. 3, pp. 127-147, 2022, doi: 10.22266/ijies2022.0630.12.

[23] P. Missier, C. McClean, J. Carlton, D. Cedrim, L. Silva, A. Garcia, A. Plastino, and A. Romanovsky, "Recruiting from the Network: Discovering Twitter Users Who Can Help Combat Zika Epidemics", In: *Proc. of Web Engineering Conference, J. Cabot, R. De Virgilio, and R. Torlone, Eds., in Lecture Notes in Computer Science*, pp. 437–445, 2017.

[24] J. E. C. Saire, "Building Intelligent Indicators to Detect Dengue Epidemics in Brazil using Social Networks", In: *Proc. of 2019 IEEE Colombian Conference on Applications in Computational Intelligence (ColCACI)*, pp. 1–5, 2019.

[25] M. Pota, M. Esposito, G. D. Pietro, and H. Fujita, "Best Practices of Convolutional Neural Networks for Question Classification", *Applied Sciences*, Vol. 10, No. 14, Art. No. 14, 2020.

[26] K. Tran, A. Bisazza, and C. Monz, "Recurrent Memory Networks for Language Modeling", *arXiv*, 2016.

[27] M. Pota, M. Ventura, H. Fujita, and M. Esposito, "Multilingual evaluation of pre-processing for BERT-based sentiment analysis of tweets", *Expert Systems with Applications*, Vol. 181, p. 115119, 2021.

[28] G. A. Pradnyana, W. Anggraeni, E. M. Yuniarno, and M. H. Purnomo, "Fine-Tuning IndoBERT Model for Big Five Personality Prediction from Indonesian Social Media", In: *Proc. of 2023 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, pp. 93–98, 2023.

[29] Q. Gu, J. Tian, X. Li, and S. Jiang, "A novel Random Forest integrated model for imbalanced data classification problem", *Knowledge-Based System*, Vol. 250, p. 109050, 2022.

[30] O. E. Gayar, T. Nasralah, and A. E. Noshokaty, "Wearable devices for health and wellbeing: Design Insights from Twitter", In: *Proc. of the 52nd Hawaii International Conference on System Sciences*, 2019.

[31] J. Oyelade, E. Uwoghiren, I. Isewon, O. Oladipupo, O. Aromolaran, and K. Michael, "Machine Learning and Sentiment Analysis: Examining the Contextual Polarity of Public Sentiment on Malaria Disease in Social Networks", In: *Proc. of BICOB 2018*, 2018.

[32] A. Khatua and A. Khatua, "Immediate and long-term effects of 2016 Zika Outbreak: A Twitter-based study", In: *Proc. of 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pp. 1–6, 2016.

[33] G. Blanco and A. Lourenço, "Optimism and pessimism analysis using deep learning on COVID-19 related twitter conversations", *Information Processing & Management*, Vol. 59, No. 3, p. 102918, 2022.

[34] C. D. A. M. Toledo, C. M. Degener, L. Vinhal, G. Coelho, W. Meira, C. T. Codeco, and M. M. Teixeira, "Dengue prediction by the web: Tweets are a useful tool for estimating and

forecasting Dengue at country and city level", *PLoS Neglected Tropical Disease*, Vol. 11, No. 7, pp. e0005729, 2017.

[35] J. A. G. Díaz, M. C. García, and R. V. García, "Ontology-driven aspect-based sentiment analysis classification: An infodemiological case study regarding infectious diseases in Latin America", *Future Generation Computer Systems*, Vol. 112, pp. 641–657, 2020.

[36] M. D. Soufi, T. S. Soltani, S. S. Vahdati, and P. R. Hachesu, "Decision support system for triage management: A hybrid approach using rule-based reasoning and fuzzy logic", *International Journal of Medical Information*, Vol. 114, pp. 35–44, 2018.

[37] J. Bedi and D. Toshniwal, "CitEnergy : A BERT based model to analyse Citizens' Energy-Tweets", *Sustainable Cities and Society*, Vol. 80, p. 103706, 2022.

[38] J. Á. González, L. F. Hurtado, and F. Pla, "TWilBert: Pre-trained deep bidirectional transformers for Spanish Twitter", *Neurocomputing*, Vol. 426, pp. 58–69, 2021.

[39] N. G. Siddappa and T. Kampalappa, "Adaptive Condensed Nearest Neighbor for Imbalance Data Classification", *International Journal of Intelligent Engineering and Systems*, Vol. 12, No. 2, pp. 104–113, 2019, doi: 10.22266/ijies2019.0430.11.

[40] T. Kim and J. S. Lee, "Maximizing AUC to learn weighted naive Bayes for imbalanced data classification", *Expert Systems with Applications*, Vol. 217, p. 119564, 2023.

[41] U. Ependi, A. F. Rochim, and A. Wibowo, "A Hybrid Sampling Approach for Improving the Classification of Imbalanced Data Using ROS and NCL Methods", *International Journal of Intelligent Engineering and Systems*, Vol. 16, No. 3, pp. 345–361, 2023, doi: 10.22266/ijies2023.0630.28.