Language Independent Models for COVID-19 Fake News Detection

Black Box versus White Box Models

W. K. Wong
Sensor Research Group, Curtin University Malaysia, Miri, Malaysia

Filbert H. Juwono
Department of Electrical and Electronic Engineering, Xi'an Jiaotong-Liverpool University, Suzhou, China

Ing Ming Chew
Department of Electrical and Computer Engineering, Curtin University Malaysia, Miri, Malaysia

Basil Andy Lease
Curtin Malaysia Research Institute, Curtin University Malaysia, Miri, Malaysia

Abstract: In an era where massive information can be spread easily through social media, fake news detention is increasingly used to prevent widespread misinformation, especially fake news regarding COVID-19. Databases have been built and machine-learning algorithms have been used to identify patterns in news content and filter the false information. A brief overview, ranging from public domain datasets through the deployment of several machine learning models, as well as feature extraction methods, is provided in this paper. As a case study, a mixed language dataset is presented. The dataset consists of tweets of COVID-19 which have been labelled as fake or real news. To perform the detection task, a classification model is implemented using language-independent features. In particular, the features offer numerical inputs that are invariant to the language type; thus, they are suitable for investigation, as many regions in the world have similar linguistic structures. Furthermore, the classification task can be performed by using black box or white box models, each having its own advantages and disadvantages. In this paper, we compare the performance of the two approaches. Simulation results show that the performance difference between black box models and white box models is not significant.

Keywords: Fake news, black box model, white box model, machine learning, COVID-19

http://doi.org/10.18080/jtde.v11n3.789

Introduction

Fake news is a term often used to describe fabricated or distorted news or stories to mislead others. The spread of fake news can have a number of negative impacts on individuals and society as a whole. Fake news can spread false information, leading to confusion and misunderstanding about important issues, such as COVID-19 (Rocha *et al.*, 2021). On the polarization of society, fake news can be used to fuel political or social division by presenting one-sided or biased information (Gupta *et al.*, 2023). Another result of fake news propagation is damage to reputation of individuals, companies, or organizations (Domenico *et al.*, 2021). Often, interference in elections is prone to be organized by unethical entities. In politics, fake news can be used to influence the outcome of elections by spreading false information about candidates or issues (Grossman & Helpman, 2023). Moreover, the spread of fake news can undermine democratic processes by spreading disinformation and sowing confusion among citizens. Looking at the above, the spread of fake news can erode trust in traditional news sources and journalism, making it harder for people to separate fact from fiction. A more devastating effect of fake news spread is the spreading of false information about health and medical issues (Waszak *et al.*, 2018). The effect of fake news spread becomes worse in the COVID-19 pandemic time, when it leads to harmful or even deadly consequences (Ferreira Caceres *et al.*, 2022).

The detection of fake news can be a challenging task, as it often involves identifying and evaluating the veracity of information that is presented as true. Some techniques for detecting fake news include fact-checking, using multiple sources to verify information, and looking for patterns of misinformation (Lin *et al.*, 2019; Zhou & Zafarani, 2019). Additionally, there are various tools and software, such as browser extensions and apps, that can help users identify fake news (Nordberga *et al.*, 2020). The most common technique used in the tools for detecting fake news is Artificial Intelligence (AI). In general, there are various AI techniques for identifying fake news, such as Natural Language Processing (NLP), machine learning, and deep learning.

Various NLP techniques can be used to automatically identify patterns in language and structure that are commonly associated with fake news (Probierz *et al.*, 2021). Machine learning techniques have been used to detect fake news by recognizing patterns and features of real and fake news and using these models to classify new, unseen news articles (Imbwaga *et al.*, 2022). Besides machine learning, deep learning techniques have also been popularly used for detecting fake news (Hu *et al.*, 2022).

The objective of this paper is to compare the performance of two model approaches in detecting fake news. We focus on COVID-19 fake news detection from Twitter news feeds (tweets) as the case study. In particular, model development, feature extraction, challenges, and public domain dataset are discussed in this paper. Moreover, we introduce a curated COVID-19 dataset of mixed Malay-English tweets, which is available on GitHub. A Support Vector Machine (SVM) is implemented as an evaluation to the separability of the classes. The rest of the paper is structured as follows. First, we discuss four fake news detection techniques. Then, challenges and remarks on the existing approaches are presented. Next, the discussion of black-box models and white-box models is presented, followed by a case study including a mixed Malay-English Twitter dataset, simulation results, and analysis. Finally, a conclusion is drawn.

Fake News Detection Techniques

In line with the objectives, this section provides some description of fake news detection strategies. According to Shu *et al.* (2017), fake news detection research works can be divided into four main categories: Data-oriented, feature-oriented, model-oriented, and application-oriented. In accordance with this division, we present the discussion in four parts. Furthermore, several of the significant problems are explored.

Data-oriented approach

The data-oriented approach in fake news detection focuses on the collection and annotation of data sets for training and testing the fake news detection models. This approach includes the process of identifying and collecting large volumes of news articles, as well as manually annotating them as real or fake. This approach aims to improve the quality and diversity of the data sets used to train and test fake news detection models. In this approach, various characteristics of the dataset, such as temporal and psychological aspects, are studied. Comprehensive datasets have been created to serve as benchmark datasets. For example, the CHECKED dataset created by Yang *et al.* (2021) includes textual, visual, temporal, and network information related to Chinese COVID-19 fake news. Melo & Figueiredo (2020) created the first Brazilian fake news dataset containing information about hashtags, media, and retweets related to COVID-19 news. Hayawi *et al.* (2022) created a dataset specifically for COVID-19 vaccine news from Twitter. Memon & Carley (2020) created a novel dataset that categorizes COVID-19 online communities into users who post misinformation and users who spread true information. Patwa *et al.* (2021) created a dataset of fake news related to COVID-19 from social media posts and articles for an online competition. Cui & Lee (2020) released the CoAID dataset, which includes fake news from social media and websites, as well as user engagement with the news. Shahi & Nandini (2020) created a fake news dataset in 40 languages related to COVID-19 and categorized the tweets into 11 categories based on the topic. Alam *et al.* (2020) released a large fake news dataset of 16,000 COVID-19 tweets in Arabic, Bulgarian, Dutch, and English languages.

In terms of temporal perspective, the spread of fake news on social media shows distinct patterns that differ from those of real news. Murayama *et al.* (2021) used two datasets of fake news items that spread on Twitter to describe the propagation of fake news as a two-stage process. Kim *et al.* (2018) tracked the time events when a story was posted to determine which story and when it would be sent for verification to fact-checkers. From a psychological standpoint, the echo-chamber effect plays an important role in capturing the intentions aspect of fake news spreading in social media. Törnberg (2018) created a simulation model to investigate the interactions between echo chambers contributing to the viral spread of misinformation on the network. Abonizio *et al.* (2020) included the sentiment polarity (negativity or positivity of a text) in their assessment, measuring the negativity or positivity of a text as part of the input features into the fake-news detection model.

Feature-oriented approach

The feature-oriented approach in fake news detection focuses on identifying and extracting relevant features from news articles that can be used to train and test fake news detection models. This includes identifying patterns in language, analysing the sentiment or tone of a text, and analysing the structure of a news article. This approach aims to improve the feature representation of news articles, which can lead to better performance of fake news detection models. According to Shu *et al.* (2017), there are two main data sources: news content and social context. For the news content data source, linguistic-based and visual-based techniques can be used to extract features from text information. Linguistic-based techniques involve extracting word features from text, which can be in either a static or dynamic form. Word embeddings represent words using vectors and are a common practice in the NLP approach. Wang *et al.* (2020) conducted a study where static representations of words, such as one-hot encoding, Bag-of-words (BoW), and Term Frequency-Inverse Document Frequency (TF-IDF), were used in the early stages of NLP. These embeddings, however, suffer from high dimensional vectors that are often as large as the vocabulary size, making them hard to use. For example, in one-hot encoding, words are represented with a one-zero vector, where all values are zero except the single value, which is one, corresponding to the word column. BoW has been used in Rusli *et al.* (2020) while Term Frequency (TF), which is similar to BoW, has been used in Jiang *et al.* (2021), and TF-IDF has been used by several authors (Hayawi *et al.*, 2022; Rusli *et al.*, 2020; Jiang *et al.*, 2021; Abdelminaam *et al.*, 2021).

Advanced static word embeddings, such as Word2Vec, have been utilized in studies by Oliveira *et al.* (2020), Ivancová *et al.* (2021) and Verma *et al.* (2021). On the other hand, Global Vectors

(GloVe) embedding has been implemented in studies by Hayawi *et al.* (2022), Jiang *et al.* (2021) and Abdelminaam *et al.* (2021). Additionally, there are fake news detection models that utilize dynamic word embeddings, such as Bidirectional Encoder Representations from Transformers (BERT), as seen in studies by Hayawi *et al.* (2022), Kar *et al.* (2020) and Hande *et al.* (2021). Other dynamic word embeddings like XLNet, Efficiently Learning an Encoder that Classifies Token Replacement Accurately (ELECTRA), and Robustly Optimized BERT Pretraining Approach (RoBERTa) have been implemented in research by Hande *et al.* (2021) and Das *et al.* (2021).

In this paper, static embedding or linguistic-based features are investigated. Static features with Language-Independent (Lang-IND) characteristics have been employed as multiple languages are investigated for fake news detection. The Lang-IND features focus on capturing high-level structures rather than specific terms from a language. In particular, linguistic-based features are extracted from text content to capture the organization of documents at different levels, including characters, words, sentences, and documents. To capture different writing styles, common lexical features are examined at the character and word level, such as total words, characters per word, frequency of large words, and unique words. Common syntactic features focus on sentence-level features, such as frequency of function words, phrases, punctuation, and part-of-speech tagging.

Sutter *et al.* (2017) used 25 language-independent features (basic frequencies of part-of-speech tags) and five language-dependent features to measure the quality of translation work from English to French and French to English done by students and compare it to the work of professionals. Abonizio *et al.* (2020) extracted language-independent features, such as complexity, stylometric, and psychological features, from textual data to detect fake news in English, Portuguese and Spanish. They found that using purely stylometric features, such as Part-of-Speech tag (POS-tag) diversity (POS-tag is a label given to each word to denote its part of speech), the ratio of named entities to text size, the ratio of quotation marks to text size, and the frequency of unrecognized words, in combination with Random Forest (RF), XG Boost, and SVM classifiers, led to an increase in model accuracy. Faustini & Covões (2020) explored features that can be used regardless of the source platform and extracted a mix of 13 features (complexity and stylometric) from news content. In addition to these features, they also extracted Word2Vec features from text and used the sum of all 100 values in the vector as the 14th feature. In the MM-COVID fake news dataset paper published in Li *et al.* (2020), features extracted from news content and social engagement patterns were described as language-invariant features for six languages (English, Spanish, Portuguese, Hindi, French and Italian). To create Lang-IND features, Veselý *et al.* (2012) trained all languages simultaneously using a Multilingual Artificial Neural Network (MANN) by modelling each language using a separate output layer. Vogel & Meghana (2020) focused mainly on features that were not tied to a specific language in order to determine fake news spreaders and attempt to block fake news from spreading at the earliest stage. They captured high-level textual features of various stylistic and psychological features, such as emojis, hashtags, upper phrases, user mentions, neutral and negative polarity.

Faustini & Covões (2019) extracted eight numerical features directly from raw texts, such as the proportion of uppercase characters, exclamation marks, question marks, text that contains exclamation marks, the number of unique words, sentences, characters and words per sentence. In addition to the eight features, the authors used POS tagging to extract the proportion of adjectives, adverbs, and nouns and three other features, including the sentiment of the message, the proportion of swear words, and the proportion of spelling errors. However, these additional features were not considered Lang-IND as they relied on tools or libraries that had been trained with a specific language to extract the features.

Some studies have combined both static and dynamic embeddings. For example, Kar *et al.* (2020) proposed a method for detecting fake news by using BERT embeddings and combined them with three stylometric features from tweet text, two user engagement features (retweet and favourite count), 19 user profile features, fact verification score, and bias score for low resource languages, such as Hindi and Bengali tweets. The study found that the feature representations extracted from Hindi and Bengali languages were highly transferable across Indic languages. Another study by Guibon *et al.* (2019) performed statistical text analysis and used a feature stacking approach on a dataset of vaccination-related fake news in English and French.

Model-oriented approach

The model-oriented approach in fake news detection focuses on developing new models for detecting fake news using machine learning algorithms, NLP techniques, and network analysis methods. This approach aims to improve the accuracy, robustness, and scalability of fake news detection models. Research conducted by Abonizio *et al.* (2020) evaluated the performance of four machine learning algorithms (*K*-Nearest Neighbours, SVM, RF, and Extreme Gradient Boosting) using Lang-IND features. Oliveira *et al.* (2020) proposed a one-class SVM model that grouped training samples into one class, with samples that did not fit into that class being placed into a new class (fake news). Kesarwani *et al.* (2020) used a *K*-Nearest Neighbour model to classify instances of fake news and found that the model achieved maximum accuracy when the value of *K* was between 15 and 20. Faustini & Covões (2020) performed a fake news detection study on multiple platforms and languages, using four machine learning algorithms, namely *K*-Nearest Neighbour, RF, Gaussian Naïve Bayes and SVM.

In the recent years, the use of deep learning classifiers has gained popularity for identifying fake news. Deep learning can be considered a subset of machine learning that uses multi-layered neural networks to build complex connections between the inputs and outputs. Research works, such as Ivancová *et al.* (2021) and Abdelminaam *et al.* (2021), have trained and compared different neural network architectures for fake news detection, such as one dimensional Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM), or modified versions of LSTM and Gated Recurrent Unit (GRU). Other research works, such as Hande *et al.* (2021) and Hayawi *et al.* (2022), proposed a fake news detection model using a deep-learning technique called Bidirectional LSTM (Bi-LSTM), which is a sequence of two LSTMs with one taking the input in a forward direction and the other in a backward direction. These models also used pre-trained transformer embeddings, such as BERT, XLNet, RoBERTa, or Glove embeddings, as input features.

Evolutionary methods, like Particle Swarm Optimization (PSO) and Salp Swarm Algorithm (SSA), have also emerged as options for reducing features for fake news detection (Al-Ahmad *et al.*, 2021). Choudhury & Acharjee (2022) proposed machine learning classifiers, such as SVM, Naïve Bayes, Logistic Regression (LR), and RF, as fitness functions in a genetic algorithm, while using TF-IDF as features input and confusion matrix to calculate evaluation metrics, such as precision, recall, and F1-score. Note that precision shows the true positive rate, while recall shows the measure of how many true positive samples the model can predict correctly out of all the true positive samples in the data. Finally, F1-score combines precision and recall, making it useful for analysing unbalanced datasets.

Application-oriented approach

The application-oriented approach in fake news detection focuses on the practical applications of fake news detection models. This includes developing systems that can automatically detect and flag fake news on social media platforms, or integrating fake-news detection models into news aggregators and search engines. This approach aims to improve the usability and effectiveness of fake news detection systems in real-world scenarios. The application-oriented approach focuses on two main areas: the diffusion of fake news; and interventions to address it. Research on fake news diffusion examines how false information spreads on social media platforms, with early studies showing that false information spreads differently than reliable information. To mitigate the effects of fake news, interventions can be implemented proactively, such as removing user accounts and labelling false news, or reactively targeting specific groups of users or the entire network, when the spread of fake news is known or unknown. Galal *et al.* (2021) suggested that reactive intervention methods involve launching campaigns to counter fake news by targeting a specific group of individuals when the affected

 http://doi.org/10.18080/jtde.v11n3.789

users are known, or targeting the entire network when the affected users are not identified. Kim *et al.* (2018) proposed frameworks for reducing the spread of fake news by flagging it for fact-checking and lowering its visibility in users' feeds.

Multilingual Fake News Detection Challenges

The main challenge in detecting fake news in languages other than English is the lack of datasets and NLP tools for those languages (De *et al.*, 2021). Research works have been done for detecting fake news in low-resource languages such as Arabic (Jardaneh *et al.*, 2019; Maakoul *et al.*, 2020), Spanish (Pizarro, 2020), Portuguese (Faustini & Covões, 2019), Indonesian (Al-Ash *et al.*, 2019; Rusli *et al.*, 2020; Prasetyo *et al.*, 2019), Slovak (Ivancová *et al.*, 2021), Chinese (Yang *et al.*, 2021), Bengali (Mugdha *et al.*, 2020) and Bangladeshi (Hussain *et al.*, 2020). However, creating datasets for these languages can be difficult and time-consuming (Kong *et al.*, 2020). Studies in Kong *et al.* (2020), Li *et al.* (2020) and Abonizio *et al.* (2020) proposed to build models that can detect fake news in multiple languages and can create a multi-language fake news dataset, which is a challenging task. Note that it can be difficult to differentiate between fake news and real news when the language is not the mother tongue (Sutter *et al.*, 2017). According to Shu *et al.* (2017), there are several characteristics of this problem that make it uniquely challenging for automated detection. One of these characteristics is that fake news is intentionally written to mislead readers, making it difficult to detect based solely on the content. Additionally, fake news can be diverse in terms of topics, styles, and media platforms, and may use diverse linguistic styles and sarcasm to distort the truth. For example, fake news may use true evidence in the wrong context to support a false claim.

Black Box vs White Box Models

It is worth noting that the above-mentioned categories are not mutually exclusive and often overlap with each other. Future research may involve combining different aspects of these categories to develop more effective fake news detection systems. In addition, as the technology and the way people consume information is changing rapidly, research in this field will have to adapt to that and keep updating the methods and techniques accordingly. This section gives a brief review of the existing developments in fake news detection from various perspectives. In particular, data acquisition, feature generation, and machine-learning models are applied accordingly. To date, there have been many research works working on Lang-IND fake news detection, such as Zervopoulos *et al.* (2022) and Imaduwage *et al.* (2022).

However, the issue with the machine learning and deep learning techniques is that they are "black box" in nature. This means that we do not understand how the decision or classification progresses. In other words, they do not offer much knowledge and transparency of how a particular set of news is labelled as legitimate or fake news. On the other hand, "white box" models are easy to interpret because they are based on patterns, rules, and decision trees (Loyola-González, 2019). Kong *et al.* (2023) proposed a two-stage evolutionary approach to generate a white box model for Lang-IND fake news detection. However, the transparency of white box models comes at the expense of reduced accuracy when compared to black box models (Fung *et al.*, 2021). It is also worth noting that the definition of lower accuracy is fuzzy. For example, improperly setting black box model parameters might result in lower accuracy than a white box model.

In this paper, we compare the performance of black box models and white box models for the Fake.my-COVID19 dataset, which is a COVID-19 bilingual Twitter dataset. We consider three black box models, i.e., SVM, K-Nearest Neighbour (KNN), and RF, and three white box models, i.e., LR, Decision Tree (DT), and Genetic Programming (GP).

Support Vector Machine (SVM)

SVM is a type of supervised learning algorithm that can be used for classification or regression tasks. The goal of an SVM is to find the best boundary (or "hyperplane") that separates the different classes in the training data. Let (x, y) be the pair of (features, label); the optimization problem in SVM is given by:

$$\min_{w,b} \|w\|^2 + c\sum_{i}^{m} \xi_i$$

subject to

$$y^{(i)}\left(w^T x^{(i)} + b\right) \geq 1 - \xi_i\,,\ i = 1, 2, \ldots, m$$

$$\xi_i \geq 0\,,\ i = 1, 2, \ldots, m$$

where w is the weights, b is bias, m is the number of data samples, ξ_i is the slack variable to allow some misclassification, and c is a penalty parameter with range between 0.5 and 1.0. Note that a larger c discourages misclassification.

SVM can also be used for non-linear classification by introducing the kernel trick, which maps the original data into a high-dimensional space where a linear boundary can be found. There are three types of kernel functions in SVM: linear, polynomial and Radial Basis Function (RBF). Using a linear kernel means that the SVM will find a linear decision boundary in the input space, which can be useful in situations where the data is linearly separable. The linear kernel is the simplest kernel function, which can be faster and use less memory compared to other more complex kernel functions, like polynomial or RBF kernel. In RBF kernel, the mapping function is given by:

$$K(x_1, x_2) = \exp\left(-\gamma \|x_1 - x_2\|^2\right)$$

where γ is a control parameter. If the γ value is small, more data samples are clustered.

K-Nearest Neighbour (KNN)

KNN is one of the most basic machine learning algorithms. It categorizes data based on a distance metric. The distance can be Euclidean distance, Minkowski distance, or Manhattan distance. It can be considered as a lazy learner since it does not "learn" until the test example is provided. As a result, whenever we have a new datum to classify, we use the training data to discover its k-nearest neighbours. Keep in mind that K is frequently an odd number to avoid ties.

Random Forest (RF)

RF is an ensemble machine learning algorithm which combines many DTs. RF can be used for classification as well as regression. A brief overview of DT will be given below. In principle, each DT will be trained on a distinct dataset, resulting in shorter depth and thereby avoiding overfitting. As a result, the DT will produce a large number of outputs. The majority of the voted classes (for classification) or the average of the individual results (for regression) will be used to make the decision.

Logistic Regression (LR)

LR can be formed by generating a multiple linear regression at the first stage, as follows:

$$z = \sum_i a_i x_i$$

where a_i are the weights and x_i are the features. The result is then input to a logistic (sigmoid) function, as follows:

$$y = \frac{1}{1+e^{-z}}$$

Decision Tree (DT)

DT is a tree-like structure that consists of a root node, decision nodes, and terminal nodes. It begins with a single node, known as the root, which reflects the initial decision. Then, branches are formed to depict the possible outcomes. Each branch leads to a new node that represents the following decision depending on the previous option. This procedure is repeated until a final result or choice is obtained, which is represented by a terminal node. Each node in the decision tree provides information that aids in deciding which branch to take, such as criteria or requirements that must be met. It is clear that the structure represents transparency, so that we can understand how the decision is taken. DT uses entropy as well as Gini entropy to split the data. There are a few algorithms that are commonly used for building a DT, such as ID3, C4.5, and Classification and Regression Trees (CART) (Javed *et al.*, 2022).

Genetic Programming (GP)

GP is an evolutionary algorithm used to find the best solution of a given problem. GP can be formed using a tree-like structure where arithmetic operators can be used as the nodes. The result is a mathematical expression that represents the solution of the problem. The arithmetic operators can be basic operators and complex operators. Similar to DT, an algorithm is used to generate the GP tree. Basically, the algorithm uses the operations similar to a Genetic Algorithm (GA): reproduction (copy the program without modifications), mutation (modify a part of the program), and crossover (two programs are selected to generate a new program).

Fake.my-COVID19 Dataset and Its Features

A public dataset is required to open the opportunity for other researchers to develop knowledge. Consider, for example, a more established domain, such as generalized optimization problems (Wong & Ming, 2019) and malware detection (Wong *et al.*, 2021), where both domains have community-based public datasets that allow for more comprehensive benchmarking, thereby enabling systematic progress. Motivated by Alameri & Mohd (2021), who encouraged work on fake news detection in Malay news, we created a Malay-English COVID-19 fake news tweets dataset that can be publicly accessed at https://github.com/z3fei/Malaysia-COVID-19-Tweet-ID/tree/main/Fake.my-COVID19. To the best of our knowledge, a fake news dataset in Malay language had not been available before we published our dataset. This has been created to contribute to the low resource Malay language. Note that bilingual mode is presented because most Malaysian people speak both English and Malay (Albury, 2017)

An initiative was undertaken to construct and refine a data collection system, which aimed at procuring COVID-19 related updates shared within Malaysia through Twitter's Standard search API. The designated timeframe for this collection spanned 1 September 2021 to 31 March 2022. Within this duration, Malaysia had initiated the administration of a third round of COVID vaccines to its healthcare frontliners and elderly people. The vaccination campaign was also extended to encompass adolescents aged between 12 and 17 years. The data collection effort yielded the accumulation of 251,216 tweets spanning a period of 231 days. From the data collected, 68% of tweets are in Malay, 28% in English, and 4% in other languages (Chinese, Tamil, etc.). We omitted the tweets containing languages other than Malay and English. In the

http://doi.org/10.18080/jtde.v11n3.789

data collection program, there are two important search criteria, namely keywords and locations (to ensure the collected tweets were the ones posted within Malaysia).

After the tweets were collected, we performed the annotation task. The process of annotation encompassed the identification of assertions put forth by users within the tweet content, subsequently cross-referencing these assertions against reliable fact-checking websites to determine their accuracy. This undertaking adhered to the guidelines outlined in Vogel & Meghana (2020) to ascertain binary classifications, distinguishing between claims classified as either fake or real. According to the guidelines, context of the tweets was considered carefully by excluding tweets that were sarcastic and/or humorous. Furthermore, the tweets that expressed general opinions regarding the vaccine, official news, and appointment details of vaccination centres were not considered as fake news. The method contributes to guaranteeing the precision of manual data annotation and upholding the elevated calibre of the dataset. Tweets that had been categorized were assigned binary labels: '1' for fake news and '0' for real news. Tweets categorized as fake typically encompassed unverifiable assertions, deceptive information, deliberate deception, or unfounded conspiracy theories lacking scientific support. To mitigate bias, the labels were initially assigned by an individual and subsequently validated by three other experts. In addition, a tweet should be marked as real news if it contained useful information on COVID-19, such as numbers, dates, vaccine progress, government policies, hotspots, etc. (Faustini & Covões, 2019). All tweets posted by government agencies, medical institutes, and official news media channels were also considered as real news (Guibon *et al.*, 2019).

Finding verifiable factual claims among the retrieved tweets was not an easy task. In Vogel & Meghana (2020), the authors suggested that tweets of less than five words should be removed. We used a filter that eliminated tweets containing fewer than 20 characters, ensuring their exclusion. Additionally, tweets comprised solely of emojis, emoticons, or greetings were disregarded. To refine the scope, a language filter was integrated to exclusively consider tweets in Malay and English, which constituted approximately 96% of the amassed tweets.

During data collection, certain keywords were often used in fake news tweets, such as *beksin* and *baksin*. The actual word for *beksin* or *baksin* is *vaksin* (Malay) or vaccine (English). The anti-vaxxers who posted these tweets misspelled the word on purpose to avoid authorities screening for the actual word (i.e., vaccine). There were other words used instead of COVID, misspelt on purpose, such as *kovid*, *kobid*, *convid*, *konvid*. Other than that, some words recorded in fake news tweets are Adverse Events Following Immunization (AEFI), ivermectin, *haram* (Eng: illegal), flu, fake, scam, deltacron, *bunuh* (Eng: kill), *cipta* (Eng: create), *racun* (Eng: poison), etc. The same user would not just post one false claim and stop spreading false claims in many weeks to come. The users that had been identified as a fake news spreader or

http://doi.org/10.18080/jtde.v11n3.789

anti-vaxxers were put into a list. All tweets posted by them would be read and further checked on their claims made against trusted sources. The Fake.my-COVID19 dataset consists of 3,068 tweets, where 1,422 tweets are fake news and the remaining tweets are real news. The extracted features are shown in Table 1.

Table 1. List of 25 Lang-IND Features

Index	Lang-IND Feature	Description
1	*cnt_sentences*	Number of sentences in tweet
2	*cnt_words*	Number of words in tweet
3	*cnt_uniquewords*	Number of unique words in tweet
4	*tweet_length*	Number of characters in tweet
5	*cnt_uniquechars*	Number of unique characters in tweet
6	*avg_words_sent*	Average number of words per sentence
7	*avg_word_length*	Average number of characters per word
8	*TTR*	Number of Type-Token Ratio (TTR) in tweet. TTR is defined as the total number of unique words divided by total words.
9	*hashtags*	Number of hashtag symbols in tweet
10	*hashtags_ratio*	Number of hashtag symbols per sentence
11	*urls*	Number of URLs in tweet
12	*urls_ratio*	Number of URLs per sentence
13	*emojis*	Number of emojis in tweet
14	*emojis_ratio*	Number of emojis per sentence
15	*puncs*	Number of punctuation marks in tweet
16	*puncs_ratio*	Number of punctuation marks per sentence
17	*exclam*	Number of exclamation marks in tweet
18	*exclam_ratio*	Number of exclamation marks per sentence
19	*question*	Number of question marks in tweet
20	*question_ratio*	Number of question marks per sentence
21	*quote*	Number of quotation marks in tweet
22	*quote_ratio*	Number of quotation marks per sentence
23	*uppercase*	Number of uppercase characters in tweet
24	*uppercase_ratio*	Number of uppercase characters per sentence
25	*alluppercase*	Number of all-uppercase words in tweet

Results and Discussion

Table 2 shows the testing mean accuracies of seven algorithms. For GP, we use two sets of functions: basic functions and extended functions. Basic functions include $\{+,-,\times,\div\}$ while, for extended functions, we use the combination of basic functions and $\{(.)^2, |.|, log1p(.), sign(.), \exp(.), \sqrt{.}\}$. The output of GP is a mathematical expression R which should be translated to the binary decision, i.e., fake or real news. The decision rule is given as follows: if R is less than 0.5, then it is real news. Otherwise, it is fake news.

Table 2. Mean Accuracy Comparison (Testing)

Algorithm	Parameters	Accuracy (mean)
Black Box Models		
SVM	Kernel = RBF $c = 0.5$ $\gamma = 1$	0.5358
	Kernel = RBF $c = 1.0$ $\gamma = 1$	0.5309
	Kernel = linear $c = 0.5$	0.8339
	Kernel = linear $c = 1.0$	0.8453
KNN	$K = 5$	0.8020
RF	Max. tree depth = 8 Max. population = 100	0.8607
White Box Models		
LR		0.8423
DT		0.8219
GP	Function = basic Depth = 4	0.8228
	Function = basic Depth = 6	0.8269
	Function = basic Depth = 8	0.8228
	Function = complex Depth = 4	0.8208
	Function = complex Depth = 6	0.8350
	Function = complex Depth = 8	0.8310
Optimized GP	Function = complex Depth = 8	0.8482

From Table 2, it can be seen that the SVMs with RBF kernel show very bad accuracy. As this is a binary classification task, we can say that the model does not learn and the decision is random. SVM with linear kernel gives better results. For the black box models, RF gives the best result of 86% accuracy. Regarding GP, it can be seen that using complex functions does not significantly improve the accuracy. Moreover, an improved accuracy can be achieved by optimizing the expression obtained from GP. It is interesting to show that we can further improve the mathematical expression R by adding some weights to the variables. The weights can be optimized by using, for example, Adaptive Differential Evolution (ADE). ADE has been used to optimize the raw GP equations as presented by Kong *et al.* (2023) and Wong *et al.* (2023). In this paper, we only optimize the GP with depth 8 and complex function.

Finally, we can see that white box models return lower accuracies compared with RF. However, we can see that the difference is not significant, i.e., less than 2%. With the transparency capability of the white box models, it would be advantageous to use white box models. Future challenges might involve further enhancement of the white box models.

Concluding Remarks

Detecting fake news in mixed languages in a limited number of characters in social media messages can be complex. In addition, there is a lack of work in fake news detection for low resource languages, such as Malay. We have created and published a Malay-English fake news Twitter dataset to support research on this topic. Furthermore, two approaches of machine learning models, i.e., black box and white box models, have been discussed and compared. Mathematical expressions can be generated by using white box models, making the model clearer to the users. Simulation results show that white box models (in terms of optimized GP) result in lower accuracy. However, the difference is insignificant; therefore, white box models with their transparency capabilities are preferred.

Acknowledgement

An earlier version of this paper was presented at ICDATE 2023, Malaysia, July 2023. We would like to express our gratitude to the editor for providing another opportunity to publish the extended version of this research work.

References

Abdelminaam, D. S., Ismail, F. H., Taha, M., Taha, A., Houssein, E. H., & Nabil, A. (2021). CoAID-DEEP: An Optimized Intelligent Framework for Automated Detecting COVID-19 Misleading Information on Twitter. *IEEE Access*, *9*, 27840–27867. https://doi.org/10.1109/ACCESS.2021.3058066

Abonizio, H. Q., Morais, J. I., Tavares, G. M., & Barbon Junior, S. (2020). Language-Independent Fake News Detection: English, Portuguese, and Spanish Mutual Features. *Future Internet*, *12*, 1–18. https://doi.org/10.3390/fi12050087

Al-Ahmad, B., Al-Zoubi, A., Abu Khurma, R., & Aljarah, I. (2021). An Evolutionary Fake News Detection Method for COVID-19 Pandemic Information. *Symmetry*, *13*, 1091. https://doi.org/10.3390/sym13061091

Alam, F., Shaar, S., Dalvi, F., Sajjad, H., Nikolov, A., Mubarak, H., Martino, G. D. S., Abdelali, A., Durrani, N., Darwish, K., Al-Homaid, A., Zaghouani, W., Caselli, T., Danoe, G., Stolk, F., Bruntink, B., & Nakov, P. (2020). Fighting the COVID-19 Infodemic: Modeling the Perspective of Journalists, Fact-Checkers, Social Media Platforms, Policy Makers, and the Society. arXiv preprint arXiv:2005.00033. https://doi.org/10.48550/arXiv.2005.00033

Alameri, S. A., & Mohd, M. (2021). Comparison of Fake News Detection Using Machine Learning and Deep Learning Techniques. 3rd International Cyber Resilience Conference (CRC). https://doi.org/10.1109/CRC50527.2021.9392458

Al-Ash, H. S., Putri, M. F., Mursanto, P., & Bustamam, A. (2019). Ensemble Learning Approach on Indonesian Fake News Classification. 3rd International Conference on Informatics and Computational Sciences (ICICoS). https://doi.org/10.1109/ICICoS48119.2019.8982409

Albury, N. J. (2017). Mother Tongues and Languaging in Malaysia: Critical Linguistics Under Critical Examination. *Language in Society*, *46,* 567–589. https://www.jstor.org/stable/26847179

Choudhury, D., & Acharjee, T. (2022). A Novel Approach to Fake News Detection in Social Networks Using Genetic Algorithm Applying Machine Learning Classifiers. *Multimedia Tools and Applications*, *82*, 9029–9045. https://doi.org/10.1007/s11042-022-12788-1

Cui, L., & Lee, D. (2020). CoAID: COVID-19 Healthcare Misinformation Dataset. arXiv preprint arXiv:2006.00885. https://doi.org/10.48550/arXiv.2006.00885

Das, S. D., Basak, A., & Dutta, S. (2021). A Heuristic-driven Ensemble Framework for COVID-19 Fake News Detection. International Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation (pp. 164–176). https://doi.org/10.48550/arXiv.2101.03545

De, A., Bandyopadhyay, D., Gain, B., & Ekbal, A. (2021). A Transformer-based Approach to Multilingual Fake News Detection in Low-resource Languages. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *21*, 1–20. https://doi.org/10.1145/3472619

Domenico, G. D., Sit, J., Ishizaka, A., & Nunan, D. (2021). Fake News, Social Media and Marketing: A Systematic Review. *Journal of Business Research*, *124*, 329–341. https://doi.org/10.1016/j.jbusres.2020.11.037

Faustini. P., & Covões, T. (2020). Fake News Detection in Multiple Platforms and Languages. *Expert Systems with Applications*, *158*, 1–17. https://doi.org/10.1016/j.eswa.2020.113503

Faustini, P., & Covões, T. (2019). Fake News Detection Using One-class Classification. 8th Brazilian Conference on Intelligent Systems (BRACIS). https://doi.org/10.1109/BRACIS.2019.00109

Ferreira Caceres, M. M., Sosa, J. P., Lawrence, J. A., Sestacovschi, C., Tidd-Johnson, A., Rasool, M. H. U., Gadamidi, V. K., Ozair, S., Pandav, K., Cuevas-Lou, C., Parrish, M., Rodriguez, I., & Fernandez, J. P. (2022). The Impact of Misinformation on the COVID-19 Pandemic. *AIMS Public Health*, *9*(2), 262–277. https://doi.org/10.3934/publichealth.2022018

Fung, P. L., Zaidan, M. A., Timonen, H., Niemi, J. V., Kousa, A., Kuula, J., Luoma, K., Tarkoma, S., Petäjä, T., Kulmala, M., & Hussein, T. (2021). Evaluation of White-box Versus Black-box Machine Learning Models in Estimating Ambient Black Carbon Concentration. *Journal of Aerosol Science*, *152*, 105694. https://doi.org/10.1016/j.jaerosci.2020.105694

Galal, S., Nagy, N., & El-Sharkawi, M. E. (2021). CNMF: A Community-Based Fake News Mitigation Framework. *Information*, *12*(9), 376. https://doi.org/10.3390/info12090376

Grossman, G. M., & Helpman, E. (2023). Electoral Competition with Fake News. *European Journal of Political Economy*, *77*, 1–12. https://doi.org/10.1016/j.ejpoleco.2022.102315

Guibon, G., Ermakova, L., Seffih, H., Firsov, A., & Noé-Bienvenu, G. (2019). Multilingual Fake News Detection with Satire. International Conference on Computational Linguistics and Intelligent Text Processing (pp. 392–402). https://doi.org/10.1007/978-3-031-24340-0_29

Gupta, M., Dennehy, D., Parra, C. M., Mäntymäki, M., & Dwivedi, Y. K. (2023). Fake News Believability: The Effects of Political Beliefs and Espoused Cultural Values. *Information & Management*. *60*, 1–12. https://doi.org/10.1016/j.im.2022.103745

Hande, A., Puranik, K., Priyadharshini, R., Thavareesan, S., & Chakravarthi, B. R. (2021). Evaluating Pretrained Transformer-based Models for COVID-19 Fake News Detection. 5th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 766–772). https://doi.org/10.1109/ICCMC51019.2021.9418446

Hayawi, K., Shahriar, S., Serhani, M. A., Taleb, I., & Mathew, S. S. (2022). ANTi-Vax: A Novel Twitter Dataset for COVID-19 Vaccine Misinformation Detection. *Public Health*, *203*, 23–30. https://doi.org/10.1016/j.puhe.2021.11.022

Hu, L., Wei, S., Zhao, Z., & Wu, B. (2022). Deep Learning for Fake News Detection: A Comprehensive Survey. *AI Open*, *3*, 133–155. https://doi.org/10.1016/j.aiopen.2022.09.001

Hussain, M G., Hasan, M. R., Rahman, M., Protim, J., & Hasan, S. A. (2020). Detection of Bangla Fake News Using MNB and SVM Classifier. arXiv preprint arXiv:2005.14627. https://doi.org/10.48550/arXiv.2005.14627

Imaduwage, S., Kumara, P. P. N. V., & Samaraweera, W. J. (2022). Importance of User Representation in Propagation Network-based Fake News Detection: A Critical Review and Potential Improvements. 2nd International Conference on Advanced Research in Computing (ICARC) (pp. 90–95). https://doi.org/10.1109/ICARC54489.2022.9754103

Imbwaga, J. L., Chittaragi, N., & Koolagudi, S. (2022). Fake News Detection Using Machine Learning Algorithms. Proceedings of the 2022 Fourteenth International Conference on Contemporary Computing (IC3-2022). https://doi.org/10.1145/3549206.3549256

Ivancová, K., Sarnovský, M., & Maslej-Krcšňáková, V. (2021). Fake News Detection in Slovak Language Using Deep Learning Techniques. IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI). http://dx.doi.org/10.1109/SAMI50585.2021.9378650

Jardaneh, G., Abdelhaq, H., Buzz, M., & Johnson, D. (2019). Classifying Arabic Tweets Based on Credibility Using Content and User Features. Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT). https://doi.org/10.1109/JEEIT.2019.8717386

Javed Mehedi Shamrat, F. M., Ranjan, R., Hasib, K. M., Yadav, A., & Siddique, A. H. (2022). Performance Evaluation Among ID3, C4.5, and CART Decision Tree Algorithm. In Ranganathan, G., Bestak, R., Palanisamy, R., & Rocha, Á. (eds). *Pervasive Computing* *and Social Networking*. *Lecture Notes in Networks and Systems*, *317*. Springer, Singapore. https://doi.org/10.1007/978-981-16-5640-8_11

Jiang, T., Li, J. P., Haq, A. U., & Saboor, A. (2020). Fake News Detection Using Deep Recurrent Neural Networks. 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP). https://doi.org/10.1109/ICCWAMTIP51612.2020.9317325

Jiang, T., Li, J. P., Haq, A. U., Saboor, A., & Ali, A. (2021). A Novel Stacking Approach for Accurate Detection of Fake News. *IEEE Access*, 9, 22626–22639. https://doi.org/10.1109/ACCESS.2021.3056079

Kar, D., Bhardwaj, M., Samanta, S., & Azad, A. P. (2020). No Rumours Please! A Multi-indic-lingual Approach for COVID Fake-tweet Detection. 2021 Grace Hopper Celebration India (GHCI) conference. https://doi.org/10.1109/GHCI50508.2021.9514012

Kesarwani, A., Chauhan, S. S., & Nair, A. R., (2020). Fake News Detection on Social Media Using K-Nearest Neighbours Classifier. International Conference on Advances in Computing and Communication Engineering (ICACCE). https://doi.org/10.1109–/ICACCE49060.2020.9154997

Kim, J., Tabibian, B., Oh, A., Schoelkopf, B., & Gomez-Rodriguez, M. (2018). Leveraging the Crowd to Detect and Reduce the Spread of Fake News and Misinformation. arXiv preprint arXiv:1711.09918. https://doi.org/10.48550/arXiv.1711.09918

Kong, J. T. H., Wong, W. K., Juwono, F. H., & Apriono, C. (2023). Generating Fake News Detection Model Using a Two-stage Evolutionary Approach. *IEEE Access*, *11*, 85067–85085. https://doi.org/10.1109/ACCESS.2023.3303321

Kong, S. H., Tan, L. M., Gan, K. H., & Samsudin, N. H. (2020). Fake News Detection Using Deep Learning. 2020 IEEE 10th Symposium on Computer Applications & Industrial Electronics (ISCAIE). https://doi.org/10.1109/DSAA49011.2020.00088

Li, Y., Jiang, B., Shu, K., & Liu, H. (2020). Mm-covid: *A Multilingual and Multimodal Data Repository for Combating COVID-19 Disinformation*. arXiv preprint arXiv:2011.04088. https://doi.org/10.48550/arXiv.2011.04088

Lin, J., Tremblay-Taylor, G., Mou, G., You, D., & Lee, K. (2019). Detecting Fake News Articles. 2019 IEEE International Conference on Big Data (Big Data) (pp. 3021–3025). http://dx.doi.org/10.1109/BigData47090.2019.9005980

Loyola-González, O. (2019). Black-Box vs. White-Box: Understanding Their Advantages and Weaknesses from A Practical Point of View. *IEEE Access*, 7, *154096–154113*. https://doi.org/10.1109/ACCESS.2019.2949286

Maakoul, O., Boucht, S., Hachimi, K., & Azzouzi, S. (2020). Towards Evaluating the COVID'19 Related Fake News Problem: Case of Morocco. 2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS). https://doi.org/10.1109/ICECOCS50124.2020.9314517

Melo, T., & Figueiredo, C. M. (2020). A First Public Dataset from Brazilian Twitter and News on COVID-19 in Portuguese. *Data in brief*, *32*, 106179. https://doi.org/10.1016–/j.dib.2020.106179

Memon, S. A., & Carley, K. M. (2020). Characterizing COVID-19 Misinformation Communities Using a Novel Twitter Dataset. arXiv preprint arXiv:2008.00791. https://doi.org/10.48550/arXiv.2008.00791

Mugdha, S. B. S., Ferdous, S. M., & Fahmin, A. (2020). Evaluating Machine Learning Algorithms for Bengali Fake News Detection. 23rd International Conference on Computer and Information Technology (ICCIT). https://doi.org/10.1109–/ICCIT51783.2020.9392662

Murayama, T., Wakamiya, S., Aramaki, E., & Kobayashi, R. (2021). Modeling the Spread of Fake News on Twitter. *PLOS ONE*, *16*(4), e0250419. https://doi.org/10.1371–/journal.pone.0250419

Nordberga, P., Kävrestada, J., & Nohlberg, M. (2020). Automatic Detection of Fake News. 6th International Workshop on Socio-Technical Perspective in IS Development (STPIS'20). https://ceur-ws.org/Vol-2789/paper23.pdf

Oliveira, N. R., Medeiros, D. S., & Mattos, D. M. (2020). A Sensitive Stylistic Approach to Identify Fake News on Social Networking. *IEEE Signal Processing Letters*, *27*, 1250–1254. http://dx.doi.org/10.1109/LSP.2020.3008087

Patwa, P., Sharma, S., Pykl, S., Guptha, V., Kumari, G., Akhtar, M. S., Ekbal, A., Das, A., & Chakraborty, T. (2021). Fighting an Infodemic: COVID-19 Fake News Dataset. In Chakraborty, T., Shu, K., Bernard, H. R., Liu, H., & Akhtar, M.S. (eds), Combating Online Hostile Posts in Regional Languages during Emergency Situation. CONSTRAINT 2021. *Communications in Computer and Information Science*, *1402*. Springer, Cham. https://doi.org/10.1007/978-3-030-73696-5_3

Pizarro, J. (2020). Profiling Bots and Fake News Spreaders at Pan'19 and Pan'20: Bots and Gender Profiling 2019, Profiling Fake News Spreaders on Twitter 2020. IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA). https://doi.org/10.1109/DSAA49011.2020.00088

Prasetyo, A., Septianto, B. D., Shidik, G. F., & Fanani, A Z. (2019). Evaluation of Feature Extraction TF-IDF in Indonesian Hoax News Classification. International Seminar on Application for Technology of Information and Communication (iSemantic). https://doi.org/10.1109/ISEMANTIC.2019.8884291

Probierz, B., Stefański, P., & Kozak, J. (2021). Rapid Detection of Fake News Based on Machine Learning Methods, *Procedia Computer Science*, *192*, 2893–2902. https://doi.org–/10.1016/j.procs.2021.09.060

Rocha, Y. M., Moura, G. A., Desidério, G. A., Oliveira C. H., Lourenço, F. D., & Figueiredo Nicolete, L. D. (2023). The Impact of Fake News on Social Media and Its Influence on Health During The COVID-19 Pandemic: A Systematic Review. *Journal of Public Health*, *31*, 1007–1016. https://doi.org/10.1007/s10389-021-01658-z

Rusli, A., Young, J. C., & Iswari, N. M. S. (2020). Identifying Fake News in Indonesian via Supervised Binary Text Classification. IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology, pp. 86–90. https://doi.org /10.1109/IAICT50021.2020.9172020

Shahi, G. K., & Nandini, D. (2020). FakeCovid –A Multilingual Cross-domain Fact Check News Dataset for COVID-19. arXiv preprint arXiv:2006.11343. https://doi.org–/10.36190/2020.14

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD explorations newsletter*, *19*(1), 22–36. https://doi.org/10.48550/arXiv.1708.01967

Sutter, G. D., Cappelle, B., Clercq, O. D., Loock, R., & Plevoets, K. (2017). Towards A Corpus-based, Statistical Approach to Translation Quality: Measuring and Visualizing Linguistic Deviance in Student Translations. *Linguistica Antverpiensia, New Series–Themes in Translation Studies*, *16*, 16–25. https://doi.org/10.52034–/lanstts.v16i0.440

Törnberg, P. (2018). Echo Chambers and Viral Misinformation: Modeling Fake News as Complex Contagion. *PLOS ONE*, *13*(9), e0203958. https://doi.org/10.1371–/journal.pone.0203958

Verma, P. K., Agrawal, P., Amorim, I., & Prodan, R. (2021). Welfare: Word Embedding Over Linguistic Features for Fake News Detection. *IEEE Transactions on Computational Social Systems*, *8*(4), 881–893. https://doi.org/10.1109/TCSS.2021.3068519

Veselý, K., Karafiát, M., Grézl, F., Janda, M., & Egorova, E. (2012). The Language-Independent Bottleneck Features. IEEE Spoken Language Technology Workshop (SLT). http://dx.doi.org/10.1109/SLT.2012.6424246

Vogel, I., & Meghana, M. (2020). Detecting Fake News Spreaders on Twitter From A Multilingual Perspective. 7th International Conference on Data Science and Advanced Analytics (DSAA). http://dx.doi.org/10.1109/DSAA49011.2020.00084

Wang, Y., Hou, Y., Che, W., & Liu, T. (2020). From Static to Dynamic Word Representations: A Survey. *International Journal of Machine Learning and Cybernetics*, *11*, 1611–1630. https://doi.org/10.1007/s13042-020-01069-8

Waszak, P., Kasprzycka-Waszak, W., Kubanek, A. (2018). The Spread of Medical Fake News in Social Media – The Pilot Quantitative Study. *Health Policy and Technology*, *7*(2), 115–118. https://doi.org/10.1016/J.HLPT.2018.03.002

Wong, W. K., Juwono, F. H., & Apriono, C. (2021). Vision-based Malware Detection: A Transfer Learning Approach Using Optimal ECOC-SVM Configuration. *IEEE Access*, *9*, 159262–159270. https://doi.org/10.1109/ACCESS.2021.3131713

Wong, W. K., Juwono, F. H., Nuwara, Y., & Kong, J. T. H. (2023). Synthesizing Missing Travel Time of P-Wave and S-Wave: A Two-Stage Evolutionary Modeling Approach. *IEEE Sensors Journal*, *23*(14), 15867–15877. http://dx.doi.org/10.1109–/JSEN.2023.3280708

Wong. W., Ming, C. I. (2019). A Review on Metaheuristic Algorithms: Recent Trends, Benchmarking and Applications. 7th International Conference on Smart Computing Communications (ICSCC). https://doi.org/10.1109/ICSCC.2019.8843624

Yang, C., Zhou, X., Zafarani, R. (2021). Checked: Chinese COVID-19 Fake News Dataset. *Social Network Analysis and Mining*, *11*(1), 1–8. https://doi.org/10.1007/s13278-021-00766-8

Zervopoulos, A., Alvanou, A. G., Bezas, K., & Papamichail, A. (2022). Deep Learning for Fake News Detection on Twitter Regarding the 2019 Hong Kong Protests. *Neural Computing and Applications*, *34*(1), 969–982. https://doi.org/10.1007/s00521-021-06230-0

Zhou, X., & Zafarani, R. (2019). Network-based Fake News Detection: A Pattern-driven Approach. *SIGKDD Explorations Newsletter*, *21*(2), 48–60. https://doi.org–/10.1145/3373464.3373473