A Meta-Analysis of State-of-the-Art Automated Fake News Detection Methods

Rafał Kozik, Aleksandra Pawlicka, Marek Pawlicki, Michał Choraś, Wojciech Mazurczyk, *Senior Member, IEEE*, and Krzysztof Cabaj

***Abstract*— Recently, various artificial intelligence (AI)-based methods have been proposed to support humans in detecting disinformation and fake news. The goal of this article is to provide a meta-analysis, and formally evaluate, compare, and benchmark various classes of fake news detection approaches. To this end, the following paper performs a comprehensive analysis of the performance-related results of different models using a range of benchmark datasets. The performed and disclosed meta-analysis compares the statistical significance of differences in a range of performance metrics, including precision, $F1$-score, recall, and balanced accuracy (BACC). The utilized approach features the 5 × 2 cross-validation methodology. The models undergoing the formal evaluation constitute state-of-the-art (SOTA) solutions meeting acceptance criteria. The evaluated approaches draw from the most recent advancements in natural language processing (NLP). The outcome of this work is the formal benchmarking and meta-analysis of fake news detection methods that can be further utilized by the research community, but more importantly by the practitioners and decision-makers that counter fake news on a daily basis, e.g., in press agencies, homeland security agencies, fact-checkers, and so on. This work is the natural extension of the authors' previous systematic analysis of fake news detection methods and authors' own fake news detection methods based on machine learning (ML)/artificial intelligence (AI) techniques.**

***Index Terms*— Detection, fake news, natural language processing (NLP), neural networks, security, text classification.**

I. Introduction

THE paths of information dissemination have witnessed dramatic shifts over recent years. A while back, the creation of news content was the domain of professionals, e.g., journalists, writers, and experts in the field, and underwent strict quality control. Before publication, the news was edited by professional editing services, employed by publishers with established reputations, who could be trusted. At every step of the process, there was the possibility to redact the piece of news, be it by the editorial boards or the authors themselves. The outcome that the readers consumed was polished, fact-checked, and adhered to certain standards [1].

Today, anyone can become a content creator and spread their views as one pleases, despite their level of expertise, education, or experience. By means of the Internet, content creators can keep their identities concealed if they wish. As a result, the opinions of random strangers, who may not have an idea what they are writing about, could appear to be worth as much as that of an expert. On top of that, once a piece of information has gone online, there is no way of preventing it from spreading. The only one-hundred-percent-sure way of stopping the information is not to publish it at all in the first place.

This reality sets the scene for the emergence of the fake news phenomenon. Although the dissemination of falsified information is most likely as old as human speech, the development of modern media has significantly changed the way deception is spread; as a consequence, it has become a highly powerful weapon. There exists no single, unequivocally agreed-upon definition of what constitutes fake news. Still, most concepts emphasize its two most significant features, namely, its authenticity and the intent behind it. Thus, fake news has basically been defined as "false stories that appear to be news, spread (· · ·) using media (Cambridge Dictionary)," but also as "intentionally written misleading content" [2], "news that is verifiably false" [2], and so on. In addition, besides conveying the meaning of fabricated content, fake news has also become an umbrella term used for other types of information, which is false and to a varying degree of maliciousness in their creation. Namely, it encompasses the whole range of false content, such as satire, false connection, misleading, fabricated, imposter or manipulated content, and false context [3]; in other words, this broad definition includes less harmful deception, such as parody, hoaxes, and rumor. Intentionally photoshopped images may fall into the category of fake news as well.

The term owes its widespread popularity to President Trump's tweets, where he used it a lot, even though he later admitted he misused the term and employed it to call out any piece of information he was not particularly fond of, no matter if it was genuine or false [4]. The general lack of consensus in understanding what fake news is has made a number of experts postulate the term to not be used at all; instead, the terms "disinformation" and "misinformation" seem to be the most often proposed ones, with the former usually meaning purposeful, malicious fabrication, and dissemination of news [5].

The key issue relating to fake news is that the content spread on the Internet poses a serious danger in real life. A piece

Manuscript received 26 December 2022; revised 5 June 2023 and 4 July 2023; accepted 13 July 2023. Date of publication 27 July 2023; date of current version 2 August 2024. This work was supported in part by the National Center for Research and Development within the INFOSTRATEG Program under Grant INFOSTRATEG-I/0019/2021-00 and in part by the National Center for Research and Development within European Interest Group (EIG) CONCERT-Japan call to the Project "Detection of fake newS on SocIal MedIa pLAtfoRms (DISSIMILAR)" under Grant EIG CONCERT-JAPAN/05/2021. *(Corresponding author: Rafał Kozik.)*

Rafał Kozik, Marek Pawlicki, and Michał Choraś are with the Telecommunications, Computer Science and Electrical Engineering, Bydgoszcz University of Science and Technology, 85-796 Bydgoszcz, Poland (e-mail: rkozik@pbs.edu.pl).

Aleksandra Pawlicka is with the Faculty of Applied Linguistics, University of Warsaw, 00-661 Warsaw, Poland.

Wojciech Mazurczyk and Krzysztof Cabaj are with the Electronics and Information Technology, Warsaw University of Technology, 00-312 Warsaw, Poland.

Digital Object Identifier 10.1109/TCSS.2023.3296627

of skillfully fabricated malicious news has the potential of ruining an individual's reputation or credibility. However, the damaging potential of deception reaches much further—to the point of threatening societies and democracy [6]. One may just wonder how the world would look today, if the result of the U.S. presidential election in 2016, allegedly influenced by massive fake news campaigns, had been different. The article which related vaccines to autism, although very quickly redacted and deemed heavily manipulated, gave rise to the anti-vaccine movement, responsible for sowing mistrust among parents, and directly blamed for the grand return of many preventable, yet potentially deadly diseases [7]. According to early estimates, hundreds of more senseless deaths may be attributed to bogus health advice that has boomed in the coronavirus pandemic, for example, the advice to drink bleach [8]. At the end of President Trump's term of office, the accusations he spread on Twitter snowballed into the Capital Riot, which took several lives. For these reasons, the burning issue of preventing the dissemination of fake news has been the subject of intense efforts.

However, once a piece of deception has been disseminated, there is not much that can be done to stop it from spreading. Instead, fake news must be combated in different ways. The two pillars of the fight against disinformation are education and detection [9]. By education, one means raising the citizens' awareness of the problem and its scale, as well as making them as vigilant as possible, so that they are not "lazy readers" who believe in everything they read, without questioning anything [10]. In turn, the detection of fake news can be realized by means of human experts, machines, or a combination of both. Although the detection manual fact-checking is the most reliable method, it is often no longer physically doable, owing to the massive flood of heterogeneous data. ML-empowered detection makes it possible to process those amounts of information, in an automated way, incomparably quicker than with human experts. The mixed approach combines the news items labeled as fake by ML tools and double-checked by human experts.

This work deals with ML-based text fake news detection. These approaches rely on various features of the texts, starting from fact-checking, to finding various telling features—linguistic and stylistic cues or the characteristics of the social context in which the news is set. ML-based methods have been proven to be effective, accurate, and powerful aids in this struggle; the deep learning (DL) methods show especially considerable promise. According to experts, DL models have the potential of detecting both the already-known kinds of deception, and the ones which are yet to be uncovered. In addition, their application requires little to no feature engineering, which, in turn, makes it possible for them to detect deception automatically, and in doing so be more effective than humans. Among the most powerful techniques, there are the methods of natural language processing (NLP), which, due to the application of the process resembling human experts' intuitive reasoning, are capable of assessing how credible the piece of information is. The DL algorithms recreate the process of intuitive reasoning by finding patterns among vast amounts of data, i.e., the stylistics, syntactic, phraseological, and morphological patterns, which are indicative of purposeful deception.

Considering the above, the goal of this article is to formally evaluate, compare, and benchmark various classes of fake news detection approaches. This work is the natural extension of the authors' previous research [11], [12] related to fake news detection methods based on machine learning (ML)/artificial intelligence (AI) techniques.

In particular, this article compares different models and demonstrates the performance-related results of these models using the existing and recently developed benchmark datasets.

The main contributions of this study are as follows.

1) To the best of authors' knowledge, it is the first analysis of the kind in the current literature.
2) The authors of this article assess four different processing paradigms, extracted on the basis of the state-of-the-art (SOTA) analysis, and formally evaluate them with the use of six industry benchmarks; this allows to provide a structured, unified comparison of the methods, providing scientifically significant contribution on which further studies can build.
3) To evaluate the existing methods, this research uses the techniques of formal cross validation, which is currently often avoided by many researchers, as cross validation significantly rises the required computational effort.
4) The obtained outcomes are further analyzed and elaborated upon.

The remainder of this article is structured as follows. In Section II, the background for the study is outlined by discussing the related works. Section III provides a detailed outline of the materials and methods, including the design of the study, the evaluated methods, the datasets, and the evaluation metrics. Section IV discusses the results of the experiments, the usefulness of obtained outcomes, and the threats to validity, followed by the final conclusions.

II. Related Research

In this section, a summary of the selected fake news detection approaches further used in the meta-analysis and benchmarking activities is contained. Hereby, the goal is not to provide an in-depth survey of all the published fake news detection methods—as we have already published a systematic mapping study in applied soft computing [13].

In the fight against fake news by means of AI, researchers apply various DL solutions; they also often leverage the specific advantages of models by joining them in various combination. This section presents some of the most noteworthy SOTA approaches to detecting deception.

As noted in [14], social media has become more significant and extensively used than ever before as a result of the COVID-19 epidemic. Unfortunately, there has also been an increase in the distribution of false information and tweets urging immediate action. In this context, the authors benchmarked several ML approaches and achieved 93% of the $F1$-score. Similarly, [15] proposed contextualized attention method, which achieved 63% of the $F1$-score on the LIAR dataset.

In [16], an overview of fake news detection approaches was presented from many viewpoints, including source-based, knowledge-based, propagation-based, style-based, and strategies based on how information spreads online. In addition, this article discusses six potential research tasks, including the detection of nontraditional fake news, to aid in the development of fake news research: early detection, identification, cross-domain analysis, explainable detection, and intervention.

Moreover, [17] provides analyses of different types of datasets concerning fake news. The authors categorize them based on the content: 1) fake news detection—classifying fake news from text and other information; 2) fact verification—judging whether a claim is true from evidence; and 3) other datasets related to fake news, which are developed neither for fake news detection nor for fact verification (e.g., fake news analysis, fake media bias analysis, or a novel task for fake news).

According to the study by Meesad [18], which compared the results of a number of methods employed to detect fake news in Thai news websites, a long short-term memory (LSTM) has proved to be the best model for the task.

In their paper, [19] has proposed tackling the issue of fake news detection by means of convolutional neural networks (CNNs), with margin loss. To address this problem, they apply various embedding models, including the static and nonstatic ones. Their best architecture, tested on two recognized datasets, information security and object technology (ISOT) and LIAR, outperforms the SOTA methods.

Aiming at achieving the highest possible accuracy of detecting fake news, [20] has designed an ensemble classification model. The model, after feature extraction, classifies the important features by means of three ML models, i.e., decision tree, random forest (RF), and extra tree classifier. The model was tested on the ISOT datasets, and the testing accuracy of the latter was higher, when compared with the SOTA.

Another approach to detecting disinformation has been the work of [21], who believes that simple classification is not enough and has to be enhanced with ML. Thus, they combined three classifiers: passive aggressive, Naïve Bayes, and support vector machine with NLP techniques. When applied on two datasets, the experimental architecture yielded encouraging results—the accuracy of up to 93%.

In their paper, [22] presents the design of a hybrid architecture, which combines a CNN with an LSTM, the purpose of which was to detect deception related to COVID-19. The solution was tested on a novel dataset created by the authors and proved highly accurate—over 98%. Similarly, [23] has combined a CNN with an LSTM, which gave very promising results on four news fake news datasets.

In turn, [24] has proposed a hybrid architecture combining a CNN with an LSTM and utilizing dimensionality reduction applied to feature vectors, which are then passed to the classifier.

The approach by Nasir et al. [25] has been to take advantage of the fact that deep neural networks specialize in certain tasks; thus, they propose a hybrid architecture, where a CNN is responsible for learning the text features, and an LSTM is to capture its sequential flow. The solution, tested on the ISOT datasets, shows promising results; the authors argue that a hybrid CNN–recurrent neural network (RNN) model is able to perform better than the baseline solutions, which do not adopt the hybrid approach.

TABLE I
SUMMARY OF RELATED WORKS INCLUDING CATEGORIZATION OF METHODS

Category of method	References
Classical Methods	[13], [16], [20], [21]
Convolutional Neural Networks	[15], [19], [22], [23], [24], [25]
LSTM-based	[18], [22], [23], [24], [25], [26]
Transformer-based	[27]

Similarly, [26] has also designed an ensemble model, combining a Bi-LSTM-GRU-dense DL model with a dense DL model. The accuracy of the proposed architecture reached 89.8%; it proves its effectiveness in detecting fake news. Moreover, it was a remarkable result in comparison with other studies on the same dataset.

Another combination of DL and NLP techniques is to be seen in [27]. The authors reported that owing to the implementation of NLP, all the models they proposed achieved the accuracy of at least 85%; the neural network architectures had the accuracy of at least 90%, and the results could be further improved, for example, if they used bidirectional encoder representations from Transformers (BERT).

In Table I, the analyzed related works are categorized according to the used processing algorithms. Classical category includes works that are not primarily based on DL techniques, but use other methods, such as ML algorithms, feature extraction, and so on. The included references are [13], [16], [20], and [21]. The CNN category includes works that mention using CNNs for fake news detection. This includes [15], [19], [22], [23], [24], and [25]. LSTM category includes works where LSTM or a combination of LSTM with other methods is used. The references included in this category are [18], [22], [23], [24], [25], and [26]. Transformers-based category includes the work that specifically mentioned the use of BERT for fake news detection, which is [27]. It is important to that some of the references are included in more than one category, as their works used a combination of methods.

III. MATERIALS AND METHODS

The aim of this work was to formally evaluate, compare, and benchmark the SOTA approaches of fake news detection and demonstrate the performance-related results of these methods using existing and recently developed benchmark datasets.

To achieve this, the SOTA fake news detection methods were identified and selected for the study. In this work, they are referred to as $M1$–$M4$. Subsequently, six recently developed, benchmark datasets were identified and included in the study. Herein, they are referred to as $D1$–$D6$.

Finally, six relevant evaluation metrics ($EM1$–$EM6$) were chosen based on the SOTA.

The design of the study was as follows: each of the selected methods ($M1$–$M4$) was evaluated on every dataset ($D1$–$D6$), and the results were measured using six evaluation metrics ($EM1$–$EM6$). The study course has been presented in Fig. 1

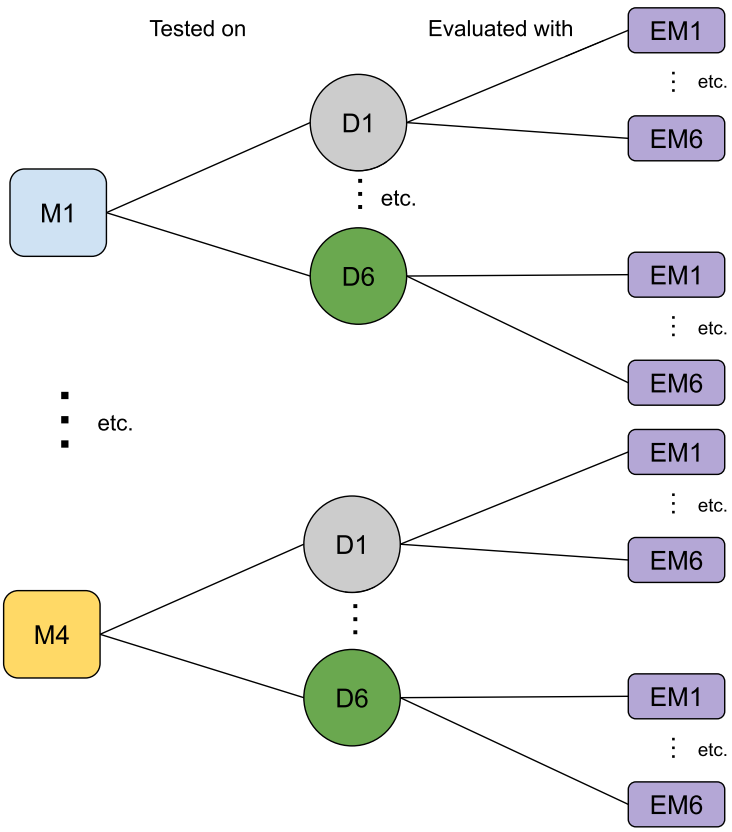


Fig. 1. Pipeline of the course of the study. Each of the chosen methods ($M1$–$M4$) was tested on each dataset ($D1$–$D6$), and the outcomes were measured using six evaluation metrics ($EM1$–$EM6$).

TABLE II
SUMMARY OF THE METHODS EVALUATED IN THIS WORK, THE DATES USED, AND THE EVALUATION METRICS APPLIED

Methods		Datasets		Evaluation metrics	
M1	CNN-based models	**D1**	MMCovid19	**EM1**	Accuracy
M2	LSTM-based models	**D2**	PubHealth	**EM2**	Balanced Accuracy
M3	BERT-based models	**D3**	Covid19FakeNews	**EM3**	F1-Score
M4	Classical models	**D4**	ISOT	**EM4**	Precision
		D5	GRAFN	**EM5**	Recall
		D6	Q-Proppy	**EM6**	G-mean

The summary of the evaluated methods, used datasets, and applied evaluation metrics has been presented in Table II.

Sections III-A–III-C present the compared methods, the datasets, and the evaluation metrics.

A. Compared Models and Methods

1) M1—CNN-Based Models: The model described in this section relies on CNN. Lately, much study focused on using CNN to various research problems (including classification and time series analysis). CNNs are oftentimes more prevalent than other models (e.g., LSTM). This is because CNN networks often run faster (e.g., when accelerated with dedicated hardware). The CNN networks have also been applied to solve particular problems (i.e., text classification) in the NLP domain. Moreover, encouraged by authors' previous research [28], we noticed that combining a CNN with classical shallow models allows us to achieve satisfactory results.

The architecture of CNN used to tackle the fake news detection problem has been presented in Fig. 2. The original text is first truncated to maximum of 500 words. Afterward, it is passed to the tokenizer, which, in this case, is fit on the available training corpus (e.g., the top-7000 most frequent words are taken). Effectively, each sentence is encoded as a 500×7000 matrix. In such a form, it is provided to the embedding layer, which transforms the sentence into a 500×32 matrix. The embedding layer is trained together with the rest of the proposed network. The embedded sentence is processed by a series of convolution and max-pooling layers. Finally, the last pooling layer is followed by a flatten layer, which arranges all the responses into a single vector that goes through one dense layer with rectified linear unit (ReLU) activation and one layer with softmax activation. The last one produces predictions for fake and true labels.

2) M2—LSTM-Based Models: The bidirectional LSTM classifier is a variant of the popular LSTM model designed to better handle longer sentences. Unlike unidirectional LSTM, it combines two LSTM models. Each network handles independently the input incoming from either positive or negative time direction. This class of models has found application in various NLP-related tasks, including fake news detection [29].

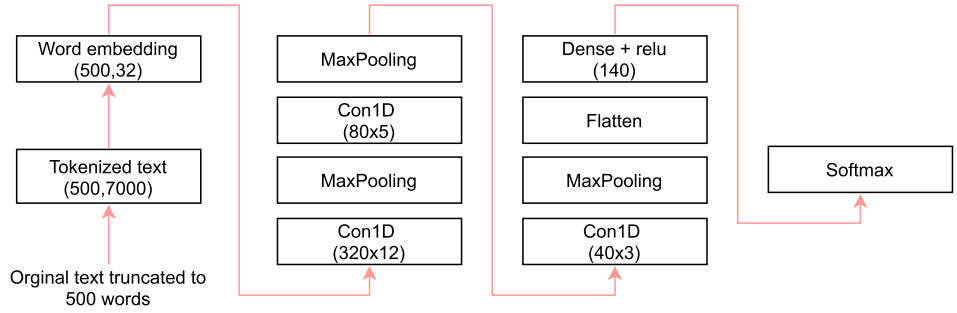


Fig. 2. CNN-based text classifier.

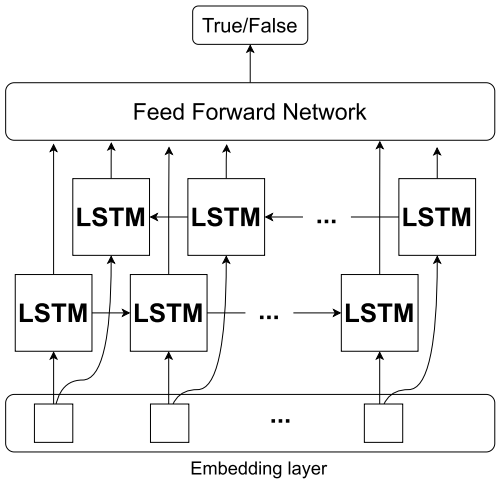


Fig. 3. LSTM-based text classifier.

The general depiction of the used bidirectional LSTM classifier is presented in Fig. 3. It comprises embedding layer, bidirectional LSTM layer, and dense output layer. The embedding layer vocabulary size is equal to 20 000, with the length of the input set to 20. It transforms the input text to 3-D tensors of the shape BATCH_SIZEx20x40. Input was first prepared by converting everything to the lower case, removing stop words, stemming, one-hot encoding, and padding.

The output from embedding layer is then used by the bidirectional LSTM layer. It contains 100 units and provides input for the final layer. The dense output layer performs the actual classification of the news using softmax activation function. Dropout through the network is set to 0.7. The model uses Adam as the optimizer, and sparse categorical cross entropy serves as the loss function.

3) M3—BERT-Based Model: This section describes architecture of the model based on the BERT [30]. This class of solutions had a notable impact on the NLP domain with performance exceeding competitors at the time. Their effectiveness is the result of the masked language model (MLM). It masks input tokens to make the model predict their ID depending on the surrounding tokens and enables them to jointly condition on both the left and right contexts. This leads to a better word representation of all the BERT approaches proposed in recent research. This comparison utilizes the DistilBERT for the use in fake news detection, as the size of the model has an impact on inference time, which is crucial from the perspective of the large-scale fake news detection system.

BERT-based approaches have already found successes in the fake news detection task. In authors' previous work, we explored the idea of matching BERT with other networks to create a hybrid solutions [31], [32]. For instance, we have found that BERT can be effectively used with the RNN. BERT is responsible for word embedding, while RNN layers on the top perform document embedding.

The architecture of the BERT-based model used during the tests is presented in Fig. 4. Each input sentence has to be first transformed by a tokenizer. It breaks them into individual tokens and then adds unique class (CLS) and separator (SEP) tokens at the beginning and the end. Finally, each token is replaced with the corresponding ID coming from the pretrained embedding table. The tokenizer was configured to either truncate or pad sentences to the maximum of 64 tokens.

The embedding layer is the complete, pretrained DistilBERT model. It is a more "lightweight" variant of BERT with reduced demand for computing resources. The training was further accelerated with the transfer learning approach [33], leaving embedding layer frozen during the process. Thus, the actual training is done only on the additional layers added on the top of the DistilBERT; bidirectional LSTM layer, pooling layer, two dense feed-forward layers, and, finally, an output layer.

The bidirectional LSTM layer has 25 units and uses hyperbolic tangent as an activation function. Recurrent activation function is set to sigmoid, while both dropout and recurrent dropout are equal to 0.6. Pooling layer is a default instance from tf.keras.layers.GlobalMaxPool1D. The dense feed-forward layers have 50 and 25 units accordingly and employ ReLU. Their dropout was set to 0.7. The output is generated by a softmax layer with only two units, classifying text as either fake or genuine. The model uses Adam as the optimizer and calculates the loss function with sparse categorical cross entropy.

4) M4—Classical Models: In authors' experiments, we have also considered classical ML solutions. According to the analysis we have conducted in Section III-A2, these approaches are still popular in the fake news detection domain. In this research, we have utilized an RF classifier. This classifier is applied on top of features extracted using pretrained word2vec embeddings. More precisely, in this article, we have utilized the RF classifier where

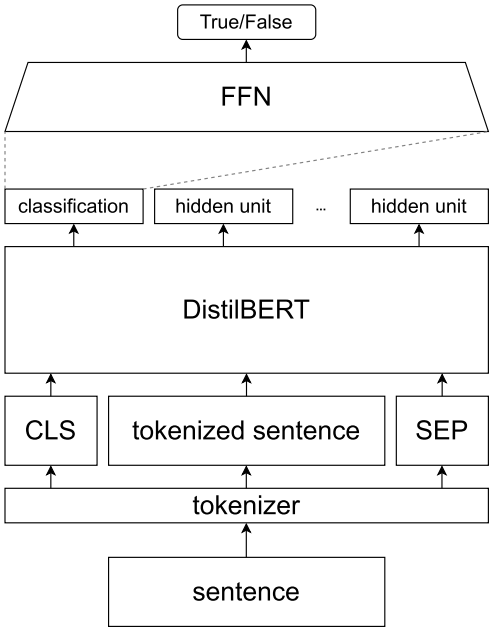


Fig. 4. BERT-based text classifier.

the number of trees (depending on the dataset) ranges from 10 to 50. Moreover, we apply average pooling to reduce the dimensionality before the classification process.

The proposed RF [34] is a different approach from an architectural point of view. It is complementary to the deep recurrent neural network described in the previous section. The main difference is that, in contrast to LSTM and Transformers-based, the RF can be trained significantly faster or even in an online manner. It means that the model can be updated right away when a new data sample is available. This substantially increases the flexibility and makes it easier to update the entire detection model when new data are available. Moreover, the reason why we focused on the RF classifier on top of embedding vectors is related to the positive results we managed to obtain for the benchmark datasets.

B. Datasets Overview

In this section, we described different datasets that have been considered to evaluate the different types of fake news detection mechanisms. The list includes the following datasets.

1) *DS1:* MMCovid19 [35]—a multilingual and multimodal data repository for combating COVID-19 disinformation. This dataset provides the multilingual fake news and the relevant social context. The authors have collected ∼4000 documents representing fake news content and ∼7000 trustworthy information from English, Spanish, Portuguese, Hindi, French, and Italian, and six different languages. In the case of the English language, there are more than 1200 fake and ∼4700 legitimate documents.
2) *DS2:* PubHealth [36]—the dataset contains documents related to public health, including biomedical subjects (e.g., infectious diseases and stem cell research) and government healthcare policy (e.g., abortion, mental health, and women's health). Each instance in the dataset has an associated label (true, false, unproven, and mixture). Moreover, each instance in the dataset has an explanation text field. The explanation is a justification for which the claim has been assigned a particular label.
3) *DS3:* Covid19FakeNews [37]—the dataset contains the list of COVID fake news/claims. It is publicly available on Zenodo repository. The dataset contains more than 9700 fake news samples and relatively few (∼500) legitimate news items.
4) *DS4:* ISOT [38] dataset contains two files: one with fake news and the other with real (true) news. The dataset contains a total of 44 898 documents, 21 417 of which are trustworthy documents and 23 481 are fake items. Each file contains four columns: article title, text, article publication date, and the subject, which can relate to one of six types of information (world news, politics news, government news, middle east, U.S. news, and left news).
5) *DS5:* GRAFN [39]—getting real about fake news is Kaggle repository dataset. The dataset contains text and metadata from 244 websites and represents almost 13 000 posts. The authors of this dataset used webhose.io application programming interface (API) to collect the data. This entire dataset is mostly dominated by political information and news from around the world.
6) *DS6:* Q-Proppy [40]—the corpus contains more than 50 000 articles from more than 100 news sources. The collected documents are labeled by the authors as either "legitimate" or "propaganda." When developing and labeling the corpus, the authors have used distant supervision technique. In that regard, articles are labeled as "propaganda" if the source (outlet/author) is considered as such by the human expert.

The distribution of documents lengths with respect to real/fake class demonstrating the variety and diversity of data contained in the benchmark datasets has been presented in Fig. 5. Moreover, in Fig. 6, diagram presents the most frequent topics tackled by the dataset. The most prominent includes coronavirus, pandemic, presidential election in United States, and general topics concerning health.

C. Evaluation Metrics

In order to compare different methods, we have used the following evaluation metrics.

1) *EM1 (Accuracy):*

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

2) *EM2 (BACC):*

$$\text{Balanced Accuracy} = \frac{\text{Recall} + \text{Precision}}{2}. \tag{2}$$

3) *EM3 (F1-Score):*

$$F\text{-Measure}(F1\text{-Score}) = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{3}$$

4) *EM4 (Precision):*

$$\text{Precision}(\text{Specificity}) = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{4}$$

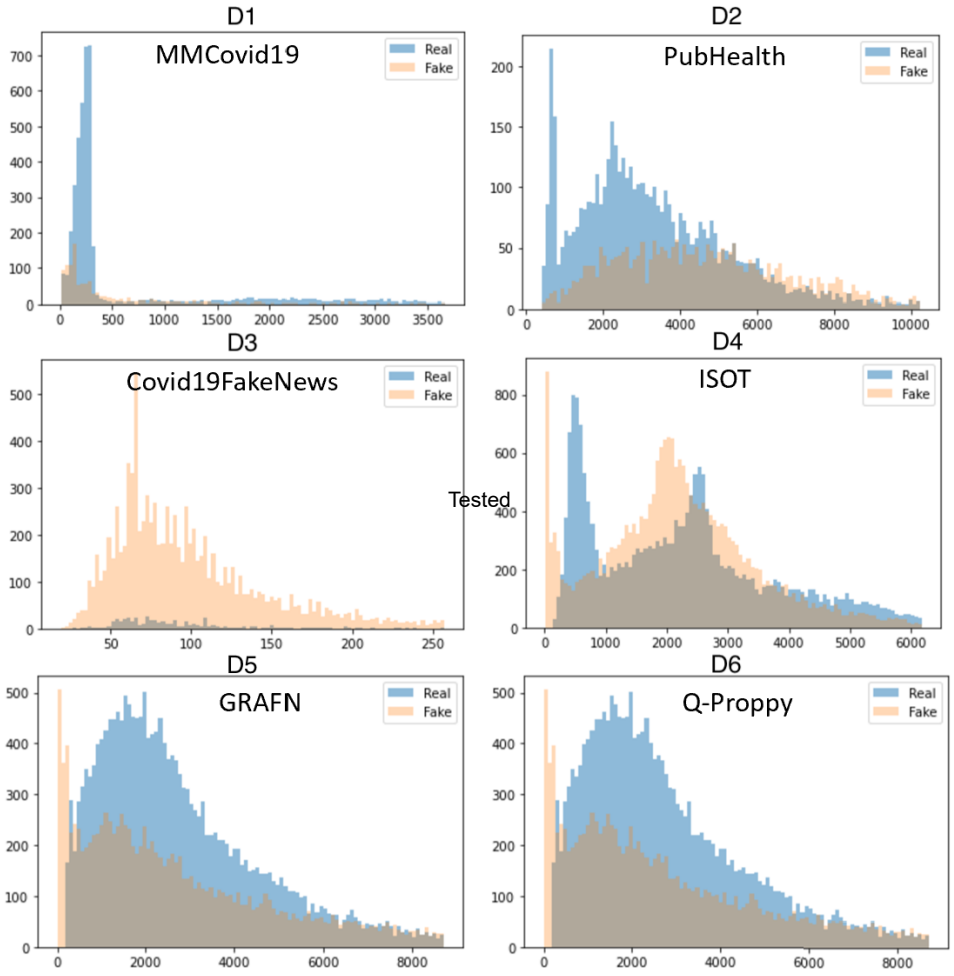


Fig. 5. Distribution of documents lengths with respect to real/fake class demonstrating the variety and diversity of data contained in the benchmark datasets.

5) *EM5 (Recall):*

$$\text{Recall(Sensitivity)} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{5}$$

6) *EM6 (G-mean):*

$$G\text{-Mean} = \sqrt{\text{Recall} * \text{Precision}}. \tag{6}$$

In the context of the metrics, TP, TN, FP, and FN indicate true positives, true negatives, false positives, and false negatives, accordingly.

IV. Results

In order to compare the metrics achieved for different algorithms, we have used the 5×2-fold cross-validation technique. In that approach, standard twofold cross validation is used, and the results are averaged. Moreover, the standard deviation from the mean value is calculated to illustrate the volatility and the significance of differences. Tables III and IV show mean values $+/-$ standard deviations for different metrics calculated for respective datasets described in Section III-B. For brevity reasons, we have chosen BACC ($EM2$) to compare the classification methods and demonstrated the differences in Fig. 7.

There are various observations, which come from the analysis of the BACC ($EM2$). First, the ISOT dataset ($D4$) seems to be the least challenging. All kinds of classification methods have reached more than 90% of BACC. Even the "classical" approaches based on TF-IDF achieve 93%. On the other hand, the dataset is well balanced in terms of class distribution (see Fig. 7), and the documents are rather long (∼2000 words per document) in their structure. Most likely, this allows the classifier to better contextualize various claims contained in the body of different documents.

Second, BERT-based approaches ($M3$) win in most of the presented cases. The exceptions are COVID-19 ($D3$) and GRAFN ($D5$). Most likely, there are two reasons why only the "classical" approaches ($M4$) achieve more than 75% of BACC ($EM2$) for COVID-19 datasets ($D3$). The first reason is that the data are heavily imbalanced, and thus, the complex models show the tendency toward the majority class and do not generalize well from the training data to unseen data. The second reason is that the documents in the $D3$ are relatively short. As a result, in most of the cases, there is no context in the document that could be leveraged by the LSTM or BERT-based approaches ($M2$ and $M3$).

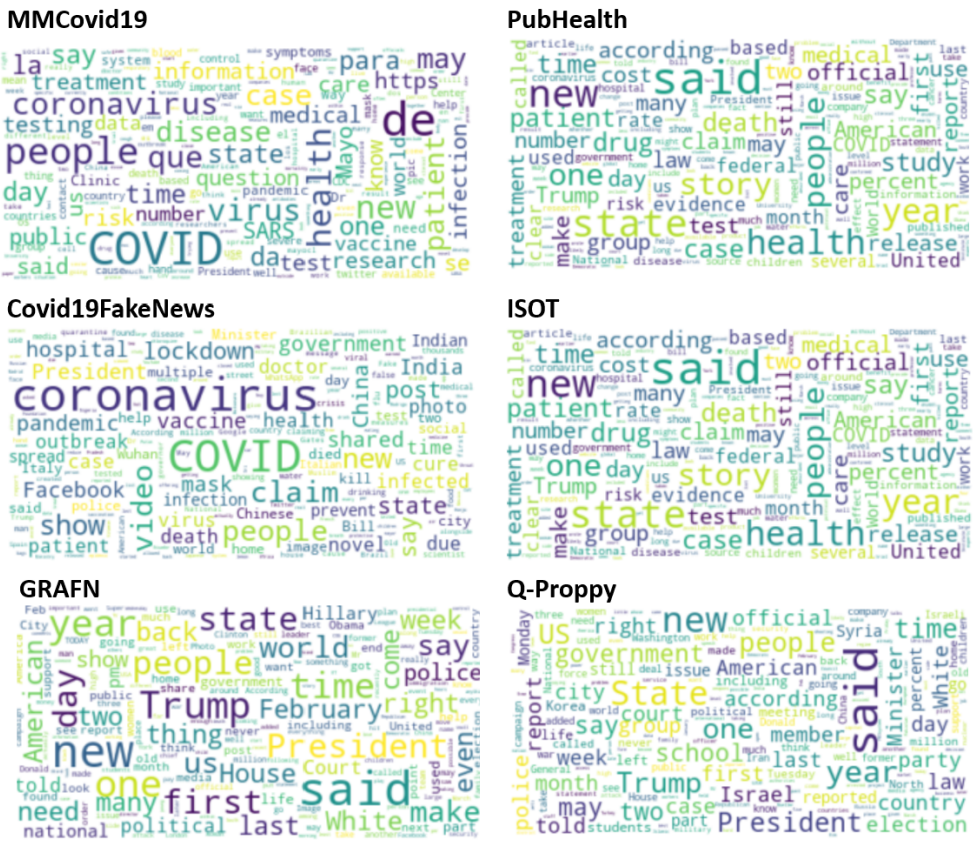


Fig. 6. Most frequent topics tackled by the document in the considered datasets. For each dataset, the diagram demonstrates the most frequent appearing topics.

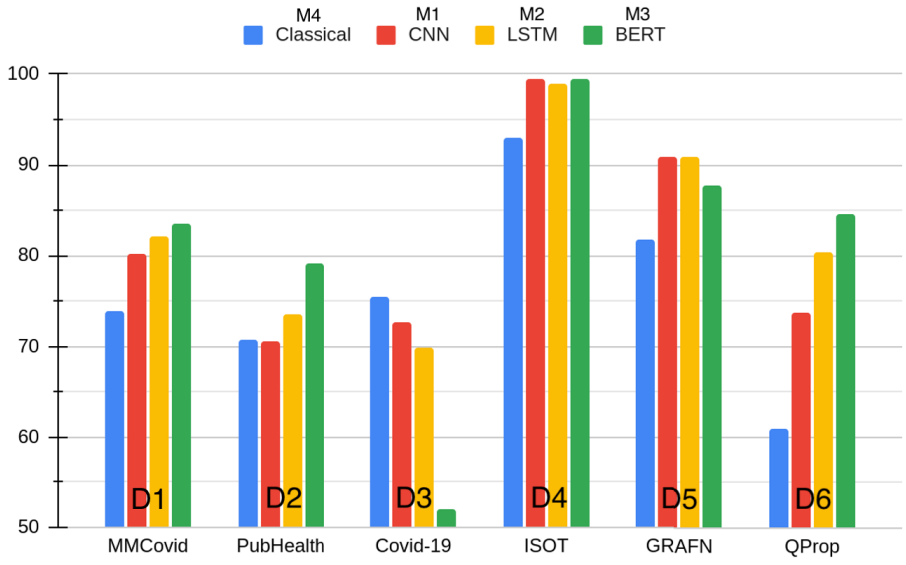


Fig. 7. Results of benchmarking of the methods included in the meta-analysis (BACC metric) with the use of varied, relevant, and novel publicly available datasets.

Finally, it may be also observed that for each of the considered datasets, there is at least one classifier, which achieves the BACC ($EM2$) higher than 75%.

A. Discussion

This article compares different models for detecting fake news. We demonstrate the performance-related results of these models using the existing benchmark datasets. In particular, we compare the statistical significance of differences in such metrics as precision, $F1$-score, recall, and BACC. In our approach, we are using the 5×2 cross-validation methodology. The compared models constitute SOTA solutions. These are applying recent algorithms and mechanisms coming from NLP domain. To the best of authors' knowledge, the current

TABLE III
COMPARISON OF VARIOUS EVALUATION METRICS (MEAN VALUE ± STANDARD DEVIATION) FOR METHODS COMPARED ON DATASETS *D*1–*D*3

Metric	M4 Classical	M1 CNN-based	M2 LSTM-based	M3 BERT-based
		DS1: MMCovid		
Accuracy	87.6 ± 0.4	88.1 ± 5.1	90.4 ± 0.6	92.5 ± 1.2
BACC	73.9 ± 0.9	80.2 ± 11.2	82.1 ± 2.1	83.6 ± 3.1
F1−score	63.3 ± 1.6	67.3 ± 23.7	75.0 ± 2.4	79.4 ± 4.2
Precision	86.8 ± 1.6	69.0 ± 25.4	85.0 ± 3.4	95.8 ± 1.3
Recall	49.8 ± 1.9	66.4 ± 23.4	67.5 ± 5.1	68.1 ± 6.5
G−mean	69.8 ± 1.3	74.6 ± 25.4	80.7 ± 2.7	82.1 ± 3.9
		DS2: PubHealth		
Accuracy	75.3 ± 0.7	70.6 ± 10.8	74.3 ± 2.2	76.9 ± 1.5
BACC	70.7 ± 0.8	70.5 ± 6.9	73.6 ± 1.2	79.1 ± 0.9
F1−score	61.4 ± 1.2	64.5 ± 5.2	67.0 ± 1.6	73.8 ± 0.9
Precision	73.4 ± 1.3	62.9 ± 10.2	65.0 ± 5.3	64.1 ± 2.3
Recall	52.7 ± 1.5	70.4 ± 14.5	70.6 ± 7.9	87.2 ± 3.2
G−mean	68.4 ± 1.0	66.6 ± 15.5	73.1 ± 1.5	78.5 ± 1.2
		DS3: Covid-19		
Accuracy	78.3 ± 0.9	96.5 ± 0.2	96.2 ± 0.6	95.5 ± 0.4
BACC	75.4 ± 1.4	72.7 ± 1.5	69.8 ± 5.6	52.0 ± 4.5
F1−score	87.3 ± 0.6	98.2 ± 0.1	98.0 ± 0.3	97.7 ± 0.2
Precision	98.3 ± 0.2	97.4 ± 0.1	97.2 ± 0.5	95.5 ± 0.4
Recall	78.6 ± 1.0	98.9 ± 0.3	98.9 ± 1.0	100.0 ± 0.0
G−mean	75.4 ± 1.5	67.7 ± 2.3	62.7 ± 9.2	11.5 ± 16.6

TABLE IV
COMPARISON OF VARIOUS EVALUATION METRICS (MEAN VALUE ± STANDARD DEVIATION) FOR METHODS COMPARED ON DATASETS *D*4–*D*6

Metric	M4 Classical	M1 CNN-based	M2 LSTM-based	M3 BERT-based
		D4: ISOT		
Accuracy	93.0 ± 0.1	99.5 ± 0.1	99.0 ± 0.3	99.4 ± 0.1
BACC	93.0 ± 0.1	99.5 ± 0.1	99.0 ± 0.3	99.4 ± 0.1
F1−score	93.3 ± 0.1	99.5 ± 0.1	99.1 ± 0.3	99.4 ± 0.1
Precision	93.6 ± 0.1	99.5 ± 0.2	99.1 ± 0.7	99.7 ± 0.1
Recall	92.9 ± 0.2	99.5 ± 0.2	99.1 ± 0.7	99.1 ± 0.1
G−mean	93.0 ± 0.1	99.5 ± 0.1	99.0 ± 0.3	99.4 ± 0.1
		D5: GRAFN		
Accuracy	84.1 ± 0.2	90.8 ± 1.1	91.4 ± 1.2	88.2 ± 0.3
BACC	81.8 ± 0.2	90.8 ± 1.0	90.9 ± 1.7	87.7 ± 0.3
F1−score	77.8 ± 0.2	88.6 ± 1.2	89.0 ± 1.8	85.1 ± 0.4
Precision	86.2 ± 0.6	87.4 ± 4.7	89.9 ± 3.4	84.9 ± 0.5
Recall	71.0 ± 0.2	90.3 ± 4.7	88.4 ± 5.3	85.2 ± 0.6
G−mean	81.1 ± 0.2	90.7 ± 1.1	90.8 ± 1.9	87.7 ± 0.3
		D6: Q-Proppy		
Accuracy	90.9 ± 0.1	92.7 ± 0.6	94.2 ± 0.5	92.7 ± 0.6
BACC	61.0 ± 0.4	73.7 ± 3.7	80.3 ± 4.1	84.6 ± 1.4
F1−score	35.5 ± 1.0	59.9 ± 5.7	70.2 ± 4.9	69.4 ± 1.5
Precision	84.9 ± 1.5	78.1 ± 3.2	82.3 ± 6.6	65.6 ± 3.8
Recall	22.5 ± 0.8	49.2 ± 7.8	62.4 ± 8.9	74.1 ± 3.6
G−mean	47.3 ± 0.8	69.3 ± 5.5	78.0 ± 5.6	83.9 ± 1.7

SOTA was missing that kind of meta-analysis. Therefore, we made the effort to fill this gap and to elaborate on the outcomes.

How can the results of the meta-analysis presented in this article be used and help in real life?

As already mentioned and proved in authors' recent work [41], fake news and disinformation can threaten democracy and harm societies. Recently, in many countries, a number of agencies and actors are created in order to counter fake news. Those can feature press agencies and tele-broadcasters, national computer emergency response teams (CERTs) and computer security incident response teams (CSIRTS), telecommunication operators and national security agencies, and nongovernmental fact-checking groups and companies. Those should, of course, include social media, such as Facebook, Meta, Twitter, and so on.

Currently, in the case of the abovementioned actors, fake news is countered manually by trained experts, annotators, and fact-checkers, which is time-consuming and expensive, and might be biased. Therefore, many of those actors start to search for automated ML-based solutions to help them indicate possible fake news and disinformation for further manual inspection (to save time and resources). The meta-analysis contained in this article (the first one ever performed, to the best of the authors' knowledge) might be helpful to decision-makers to establish relevant processes and use such automated methods in practice in order to effectively counter a huge societal challenge, namely, disinformation.

B. Threats to Validity

This work performs a formal comparison of a range of state-of-the-art NLP approaches to the detection of disinformation. However, there is a range of threats to validity embedded a priori in these approaches. First, as the NLP methods only consider the specific wording of a piece of text, the usefulness of certain approaches might change when tested on different languages and in different domains. As such, the NLP-based methods for the handling of fake news require a significant overhead in the creation of adequate corpora and annotation of reliable and unreliable pieces of information to feed and train the algorithms. It is yet to be measured how fast the way fake news is written changes and how often the algorithms would need to be updated. Second, the detection results, while very accurate in certain domains, do not really provide an insight into what the detection is based on. This erodes the societal trust in such approaches. Explainability for AI (xAI) techniques should be provided and employed to at least peek into the black box. Third, all AI-based algorithms are susceptible to the bias embedded in the data they were trained with. Utmost caution and vigilance should be employed while annotating the training data, preferably with the inclusion of fairness-maintaining algorithms to prevent bias. Finally, the nefarious actors who spread fake news could use the state-of-the-art NLP-based methods trained on open benchmark datasets to reverse-engineer the detection process and use the inferred knowledge to write pieces of disinformation that are even harder to spot in an automated way, or are able to evade detection entirely [12].

V. CONCLUSION

This article presented the results of a study the goal, which was to formally evaluate, compare, and benchmark various classes of SOTA fake news detection approaches. This was done using recently developed, benchmark datasets. Then, the performance-related results of the models were

presented, making this article the first comprehensive, complex meta-analysis of the fake news detection models. This article methodologically evaluates four existing methods and approaches, performing a formal cross-validation experiment (which are often avoided by many researchers due to the effort necessary), assessing the four different processing paradigms extracted on the basis of the SOTA analysis and formally evaluating them with the use of six industry benchmarks. This allows the article to provide a structured, unified comparison of the methods, providing a scientifically significant contribution. The results are a valuable resource for decision-makers to establish relevant processes and use automated fake news detection methods in practice. Moreover, it is worth to add that the analyzed methods and the results of our methodological meta-analysis and comparison can be used for various applications (also homeland security) and are valid for both debunking and prebunking approaches to counter disinformation.

Currently, conducted end described in this article experiments utilize benchmark datasets. As future work, we plan to integrate the best detection algorithms into the DISSIMILAR harvester system [42]. In effect, any user of this system will be able to check if the provided URL contains fake news.

REFERENCES

[1] Z. Turk, "Technology as enabler of fake news and a potential tool to combat it," Eur. Parliament, Strasbourg, France, Tech. Rep. 619008, 2018.

[2] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *J. Econ. Perspect.*, vol. 31, no. 2, pp. 211–236, May 2017. [Online]. Available: https://pubs.aeaweb.org/doi/10.1257/jep.31.2.211

[3] C. Wardle, "Fake news. It's complicated," *First Draft*, 2017. [Online]. Available: https://firstdraftnews.org/articles/fake-news-complicated/

[4] D. Lind, "President Donald Trump finally admits that 'fake news' just means news he doesn't like," *Vox*, 2018. [Online]. Available: https://www.vox.com/policy-and-politics/2018/5/9/17335306/trump-tweet-twitter-latest-fake-news-credentials

[5] F. Giuliani-Hoffman. (2017). F*** news should be replaced by these words, Claire Wardle says. CNN Business. [Online]. Available: https://money.cnn.com/2017/11/03/media/claire-wardle-fake-news-reliable-sources-podcast/index.html

[6] L. Curtin, "The 'fake' news effect," *Amer. Nurse*, 2020. [Online]. Available: https://www.myamericannurse.com/the-fake-news-effect/

[7] J. Belluz, "Research fraud catalyzed the anti-vaccination movement. Let's not repeat history," *Vox*, 2019. [Online]. Available: https://www.vox.com/2018/2/27/17057990/andrew-wakefield-vaccines-autism-study

[8] M. S. Islam et al., "COVID-19–related infodemic and its impact on public health: A global social media analysis," *Amer. J. Tropical Med. Hygiene*, vol. 103, no. 4, pp. 1621–1629, Oct. 2020. [Online]. Available: https://ajtmh.org/doi/10.4269/ajtmh.20-0812

[9] M. Taboada, "Authentic language in fake news," *Items*, 2021. [Online]. Available: https://items.ssrc.org/beyond-disinformation/authentic-language-in-fake-news/

[10] G. Pennycook and D. G. Rand, "Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning," *Cognition*, vol. 188, pp. 39–50, Jul. 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S001002771830163X

[11] P. Ksieniewicz, P. Zyblewski, W. Borek-Marciniec, R. Kozik, M. Choraś, and M. Woźniak, "Alphabet flatting as a variant of n-gram feature extraction method in ensemble classification of fake news," *Eng. Appl. Artif. Intell.*, vol. 120, Apr. 2023, Art. no. 105882.

[12] M. Szczepański, M. Pawlicki, R. Kozik, and M. Choraś, "New explainability method for BERT-based model in fake news detection," *Sci. Rep.*, vol. 11, no. 1, p. 23705, Dec. 2021, doi: 10.1038/s41598-021-03100-6.

[13] M. Choraś et al., "Advanced machine learning techniques for fake news (online disinformation) detection: A systematic mapping study," *Appl. Soft Comput.*, vol. 101, Mar. 2021, Art. no. 107050.

[14] P. Patwa et al., "Fighting an infodemic: COVID-19 fake news dataset," in *Combating Online Hostile Posts in Regional Languages During Emergency Situation*. Switzerland: Springer, 2021, pp. 21–29.

[15] E. Ranjan, "Fake news detection by learning convolution filters through contextualized attention," ResearchGate, GitHub, Tech. Rep., Aug. 2019. [Online]. Available: https://github.com/ekagra-ranjan/fake-news-detection-LIAR-pytorch

[16] X. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," *ACM Comput. Surv.*, vol. 53, no. 5, pp. 1–40, Sep. 2020.

[17] T. Murayama, "Dataset of fake news detection and fact verification: A survey," 2021.

[18] P. Meesad, "Thai fake news detection based on information retrieval, natural language processing and machine learning," *Social Netw. Comput. Sci.*, vol. 2, no. 6, p. 425, Nov. 2021. [Online]. Available: https://link.springer.com/10.1007/s42979-021-00775-6

[19] M. H. Goldani, R. Safabakhsh, and S. Momtazi, "Convolutional neural network with margin loss for fake news detection," *Inf. Process. Manage.*, vol. 58, no. 1, Jan. 2021, Art. no. 102418. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0306457320309134

[20] S. Hakak, M. Alazab, S. Khan, T. R. Gadekallu, P. K. R. Maddikunta, and W. Z. Khan, "An ensemble machine learning approach through effective feature extraction to classify fake news," *Future Gener. Comput. Syst.*, vol. 117, pp. 47–58, Apr. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167739X20330466

[21] S. Ahmed, K. Hinkelmann, and F. Corradini, "Development of fake news model using machine learning through natural language processing," 2022, *arXiv:2201.07489*.

[22] R. Kaliyar, A. Goswami, and P. Narang, "A hybrid model for effective fake news detection with a novel COVID-19 dataset," in *Proc. 13th Int. Conf. Agents Artif. Intell.* Setúbal, Portugal: SciTePress, 2021, pp. 1066–1072. [Online]. Available: https://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0010316010661072

[23] K. L. Tan, C. Poo Lee, and K. M. Lim, "Fake news detection with hybrid CNN-LSTM," in *Proc. 9th Int. Conf. Inf. Commun. Technol. (ICoICT)*, Aug. 2021, pp. 606–610. [Online]. Available: https://ieeexplore.ieee.org/document/9527469/

[24] M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi, and B.-W. On, "Fake news stance detection using deep learning architecture (CNN-LSTM)," *IEEE Access*, vol. 8, pp. 156695–156706, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9178321/

[25] J. A. Nasir, O. S. Khan, and I. Varlamis, "Fake news detection: A hybrid CNN-RNN based deep learning approach," *Int. J. Inf. Manage. Data Insights*, vol. 1, no. 1, Apr. 2021, Art. no. 100007. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2667096820300070

[26] N. Aslam, I. U. Khan, F. S. Alotaibi, L. A. Aldaej, and A. K. Aldubaikil, "Fake detect: A deep learning ensemble model for fake news detection," *Complexity*, vol. 2021, pp. 1–8, Apr. 2021. [Online]. Available: https://www.hindawi.com/journals/complexity/2021/5557784/

[27] C.-M. Lai, M.-H. Chen, E. Kristiani, V. K. Verma, and C.-T. Yang, "Fake news classification based on content level features," *Appl. Sci.*, vol. 12, no. 3, p. 1116, Jan. 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/3/1116

[28] M. Choraś, M. Pawlicki, R. Kozik, K. Demestichas, P. Kosmides, and M. Gupta, "SocialTruth project approach to online disinformation (fake news) detection and mitigation," in *Proc. 14th Int. Conf. Availability, Rel. Secur.* New York, NY, USA: Association for Computing Machinery, Aug. 2019, pp. 1–10, doi: 10.1145/3339252.3341497.

[29] P. Bahad, P. Saxena, and R. Kamal, "Fake news detection using bi-directional LSTM-recurrent neural network," *Proc. Comput. Sci.*, vol. 165, pp. 74–82, Jan. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050920300806

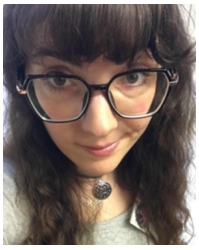
[30] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, vol. 1. Minneapolis, MN, USA: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. [Online]. Available: https://aclanthology.org/N19-1423, doi: 10.18653/v1/N19-1423.

[31] S. Kula, M. Choraś, and R. Kozik, "Application of the BERT-based architecture in fake news detection," in *Proc. 13th Int. Conf. Comput. Intell. Secur. Inf. Syst. (CISIS)*, Á. Herrero, C. Cambra, D. Urda, J. Sedano, H. Quintián, and E. Corchado, Eds. Cham, Switzerland: Springer, 2021, pp. 239–249.

[32] S. Kula, R. Kozik, and M. Choraś, "Implementation of the BERT-derived architectures to tackle disinformation challenges," *Neural Comput. Appl.*, vol. 34, no. 23, pp. 20449–20461, Dec. 2022.
[33] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[34] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
[35] Y. Li, B. Jiang, K. Shu, and H. Liu, "MM-COVID: A multilingual and multimodal data repository for combating COVID-19 disinformation," 2020, *arXiv:2011.04088*.
[36] N. Kotonya and F. Toni, "Explainable automated fact-checking for public health claims," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*. Cedarville, OH, USA: Association for Computational Linguistics, Nov. 2020, pp. 7740–7754. [Online]. Available: https://www.aclweb.org/anthology/2020.emnlp-main.623
[37] S. Banik, "COVID fake news dataset [data set]," Zenodo, Tech. Rep., 2020, doi: 10.5281/zenodo.428252.
[38] H. Ahmed, I. Traore, and S. Saad, "Detecting opinion spams and fake news using text classification," *Secur. Privacy*, vol. 1, no. 1, p. e9, Jan. 2018. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/spy2.9
[39] M. Risdal. (2016). Getting real about fake news. Kaggle. [Online]. Available: https://www.kaggle.com/mrisdal/fake-news
[40] A. Barrón-Cedeño, I. Jaradat, G. Da San Martino, and P. Nakov, "Proppy: Organizing the news based on their propagandistic content," *Inf. Process. Manage.*, vol. 56, no. 5, pp. 1849–1864, Sep. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306457318306058
[41] M. Choraś, A. Pawlicka, R. Kozik, and M. Woźniak, "How machine learning may prevent the breakdown of democracy by contributing to fake news detection," *IT Prof.*, vol. 24, no. 2, pp. 25–31, Mar. 2022.
[42] D. Megías, M. Kuribayashi, A. Rosales, K. Cabaj, and W. Mazurczyk, "Architecture of a fake news detection system combining digital watermarking, signal processing, and machine learning," *J. Wireless Mobile Netw., Ubiquitous Comput., Dependable Appl.*, vol. 13, no. 1, pp. 33–55, Mar. 2022.

Rafał Kozik received the Ph.D. degree in telecommunications engineering from the University of Science and Technology (UTP), Bydgoszcz, Poland, in 2013, and the D.Sc. degree in computer science engineering from the West Pomeranian University of Technology, Szczecin, Poland, in 2019.

He is currently a Professor with the Bydgoszcz University of Science and Technology, Bydgoszcz. Since 2009, he has been involved in a number of international and national research projects related to cybersecurity, critical infrastructures protection, software quality, and data privacy (e.g., FP7 INfrastructure for heTErogeneous, Resilient, SEcure, Complex, Tightly Inter-Operating Networks (INTERSECTION), FP7 INcreasing Security and Protection through Infrastructure REsilience (INSPIRE), FP7 Comprehensive Approach to cyber roadMap coordINation and develOpment (CAMINO), FP7 Critical Infrastructure Preparedness and Resilience Research Network (CIPRNet), SOPAS, SECOR, and H2020 Q-Rapids). He has authored over 140 reviewed scientific publications.



Aleksandra Pawlicka received the Ph.D. degree from Nicolaus Copernicus University, Toruń, Poland, in 2014.

She is a philologist and a research and development specialist. She is interested in computer science, linguistics, language teaching and learning, and pedagogy; in her works, she combines those fields. She has authored a number of multidisciplinary scientific publications and has been involved in several international projects, such as H2020 secure intelligent methods for advanced recognition of malware and stegomalware (SIMARGL), H2020 Special projects for advanced research and technology in Europe (SPARTA), and H2020 Prediction and Visual Intelligence for Security Information (PREVISION).

Michał Choraś is currently a Full Professor (title granted in 2021) with the Bydgoszcz University of Science and Technology, Bydgoszcz, Poland, where he is the Head of the Teleinformatics Systems Division and the Pattern Recognition and Applied Security (PATRAS), Research Group. He is affiliated also with FernUniversität, Hagen, Germany, where he was a Project Coordinator for H2020 secure intelligent methods for advanced recognition of malware and stegomalware (SIMARGL). He has authored over 280 reviewed scientific publications. He has been involved in many European Union (EU) projects (e.g., SocialTruth, FP7 Critical Infrastructure Preparedness and Resilience Research Network (CIPRNet), Q-Rapids, and InfraStress). His research interests include data science, artificial intelligence (AI), and pattern recognition in several domains, e.g., cyber security, image processing, software engineering, prediction, anomaly detection, correlation, biometrics, and critical infrastructures protection.

Marek Pawlicki received the Ph.D. degree in engineering from the Bydgoszcz University of Science and Technology, Bydgoszcz, Poland, in 2020.

He is currently an Adjunct Professor with the Bydgoszcz University of Science and Technology, Bydgoszcz, Poland. He has been involved in a number of international projects related to cybersecurity, critical infrastructures protection, and software quality (e.g., H2020 Special projects for advanced research and technology in Europe (SPARTA), H2020 secure intelligent methods for advanced recognition of malware and stegomalware (SIMARGL), H2020 Prediction and Visual Intelligence for Security Information (PREVISION), H2020 Multimedia Analysis and Correlation Engine for Organised Crime Prevention and Investigation (MAGNETO), H2020 Q-Rapids, and H2020 SocialTruth). He has authored over 60 peer-reviewed scientific publications. His research interests include the application of machine learning in several domains, including cybersecurity.

Wojciech Mazurczyk (Senior Member, IEEE) received the B.Sc., M.Sc., Ph.D. (Hons.), and D.Sc. (Habilitation) degrees in telecommunications from the Warsaw University of Technology (WUT), Warsaw, Poland, in 2003, 2004, 2009, and 2014, respectively.

He is currently a University Professor with the Institute of Computer Science, WUT, where he is the Head of the Computer Systems Security Group. He also works as a Researcher with the Parallelism and VLSI Group, Faculty of Mathematics and Computer Science, FernUniversität, Hagen, Germany. His research interests include bioinspired cybersecurity and networking, information hiding, and network security.

Dr. Mazurczyk is involved in the technical program committee of many international conferences. Between 2018 and 2021, he served as an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY. He serves as a reviewer for major international magazines and journals. Since 2016, he has been an Editor-in-Chief of an open access *Journal of Cyber Security and Mobility*.

Krzysztof Cabaj received the M.Sc., Ph.D., and D.Sc. (Habilitation) degrees in computer science from the Faculty of Electronics and Information Technology, Warsaw University of Technology (WUT), Warsaw, Poland, in 2004, 2009, and 2019, respectively.

He was an Instructor of Cisco certificated academy courses, such as Cisco Certified Network Associate (CCNA) Routing and Switching, CCNA Security, and Cisco Certified Network Professional (CCNP), at the International Telecommunication Union Internet Training Centre (ITU-ITC). He is currently a University Professor with the Institute of Computer Science, WUT, where he is also a Coleader of the Computer Systems Security Group, Institute of Computer Science. He has authored or coauthored over 70 publications, and is a supervisor of more than 25 M.Sc. and B.Sc. degree theses in the field of information security. He took part in over a dozen research projects, among others for European Union (EU), European Space Agency (ESA), Samsung, U.S. Army, and U.S. Air Force. His research interests include network security, honeypots, dynamic malware analysis, data-mining techniques, the Internet of Things (IoT), and industrial control systems security.