Detection of Fake News Using Deep Learning and Machine Learning

Gabriela CHIRIAC, Ada Maria CATINA
Bucharest University of Economic Studies
Faculty of Cybernetics, Statistics and Economic Informatics
Bucharest, Romania
chiriacgabriela20@stud.ase.ro; catinaada20@stud.ase.ro

Automatically identifying fake news is a complex challenge, involving detailed knowledge of how fake news is propagated and advanced data processing technologies. The use of Machine Learning and Deep Learning algorithms for misinformation detection involves a continuous learning process as manipulation methods are constantly evolving. Effective detection of fake news requires constant adaptation of algorithms to keep up with new disinformation methods. While these technologies offer promising solutions, the challenge is to calibrate them properly so that they work optimally in different contexts.
This paper explores automated methods for detecting fake news, analyzing the effectiveness of the various techniques and how they can be improved to face the challenges of data quality, domain variability and the continuous evolution of disinformation strategies.

Keywords: *misinformation, Machine Learning, Deep Learning, fake news.*

1 Introduction

In the age of digitalization and social media platforms, the spread of fake news has become a real challenge. Today, online platforms provide us with access to information, but this freedom also comes with the risk of falling prey to cybercriminals who spread fake news.

Fake news is defined as text, images, videos or audio files intentionally created to misinform public opinion. Social media platforms enable misinformation to spread very quick, exposing thousands of users just in few minutes. Their purpose is to influence people's opinions about an individual or group of individuals, to create social unrest and to undermine trust in institutions.

Because of the quick spread and big volume of information, distinguishing between real and false news has become very difficult. Manually verifying information can be a slow and difficult process due to the large amount of data being produced every day.

Social media is a key way to consume news, but it can have both advantages and disadvantages. It is great for access, quick and cheap, as well, but it also makes the spread of fake news easier. In many cases, the quality of information can be unreliable, as some information are intentionally falsified. Detecting this type of deception has become a research topic, attracting increased interest in recent years [1]. According to one of the EU Agency for Cybersecurity reports [2], disinformation and misinformation are a critical challange to global security. In this report we can find that the use of cloud computing technologies and AI algorithms contributes to the creation of misinformation, accentuating challenges related to the *security and integrity of the online environment*.

Another important aspect highlighted by the European data Protection Supervisor [3] is the use of complex applications called *bots*, which are organized into networks and used heavily to amplify disinformation content. Bot networks are organized by foreign actors attempting to

mask the true initiators, thus further complicating the detection process. However, recent research indicates that while automated bots contribute to the spread of fake news, the influence of human behavior, especially *word of mouth* marketing, has an even greater impact. As such, the issue of how to combat the spread of fake news can't come down to how to identify the people who produce and spread it. We must equip people to spot fake news and teach them critical thinking so they can keep the danger away. Thus, media education becomes the key factor in the fight against fake news, helping to build a society that is better prepared to deal with manipulation [3].

This phenomenon has serious consequences for society, and various advanced methods have been developed to combat it, applying machine learning and deep learning algorithms. The algorithms are trained to verify news content by exposing them to large volumes of data. This training process involves the algorithms learning to distinguish between real and fake news based on a set of properties extracted from the text, images, or metadata associated with the news. Since the spread of fake news is constantly changed, these methods need to continuous update to keep up with sophisticated tactics used by disinformation creators to mislead both users and detection systems. [3]

This paper aims to explore advanced *Machine Learning and Deep Learning* methods used in the automatic detection of fake news. This study will examine different techniques and algorithms that can be applied to classify informational content, aiming to evaluate their effectiveness in detecting fake news and combating them.

2. Literature Review

Numerous studies have explored solutions for identifying and classifying fake news from various data sources, using different approaches to increase the accuracy of predictions. In this section, we will analyze significant contributions from the research literature that address various techniques and methods for detecting this problem.

In 2015, several researchers [4] adopted a text-based approach, dividing fake news into three typologies: serious inventions, large-scale jokes, and parodies. They explore the difficulties encountered in developing a false news detection system, highlighting the importance of taking into account the typology and context in which they are generated. Perez-Rosas and other researchers [5] have developed an automated algorithm that combines lexical, syntactic and semantic information to detect fake news, bringing more complexity and efficiency to the detection process. In a similar study [6] proposed a hybrid method for detecting fake news, which combines linguistic analysis with social networks evaluation techniques. This method has proven to be successful in identifying fake news, given that social network analysis can reveal the ways in which false information is propagated and amplified on online platforms.

In 2016, Dadgar and collaborators [7] applied feature-extracting techniques such as TF-IDF, and implemented machine learning algorithms such as SVM (support Vector machines) to classify fake news into various categories. Ruchansky, Seo, and Liu [8] proposed a hybrid algorithm named CSI, which integrates three essential components: capture, score, and integration to improve predictions about fake news. This approach underlines the importance of integrating information from multiple sources to build a more robust detection system.

Regarding social network-based techniques, Shu, Wang and Liu [9] studied another model that analyzes factors such as the position of news and user interactions on social networks. This approach is extremely relevant, given how much fake

news and manipulation rely on social media platforms. Similarly, a study by Jin, Cao, Zhang, and Luo [10] proposed an approach based on identifying conflicting points of view in social networks. By applying this method to real data sets, the authors demonstrated how differences of opinion can be indicators of the presence of misleading information.

Several researchers participating in a conference on natural language processing [11] used an attention-based long short-term memory (LSTM) network to detect fake news, demonstrating the use of advanced technologies, such as neural networks, to improve the detection process.

Recent studies, such as that of Janze Christian [12], have applied these techniques in the context of the 2016 US election to observe the impact of fake news on election campaigns. Butain and Golbeck [13] have created an automated system that detects fake news on Twitter, a platform where its dissemination is a major problem.

To reduce the impact of fake news, several researchers [14] have proposed a competitive model that analyzes the relationship between the original and updated false information, with the aim of minimizing its effect on the public.

On the other hand, Tschiatschek and collaborators [15] proposed a method based on the human signals, using Bayesian inference to improve detection accuracy, and Guacho, Abdali and Papalexakis [16] introduced a semi-supervised false news detection technique, which combines human signals with machine learning algorithms.

These innovative research and solutions highlight the complexity and diversity of approaches in the field of detecting fake news. While technologies have advanced in detection of fake news, the challenges remain due to the quick spread of fake news and to the development of new methods and tools used for disinformation.

3. Methodology

This paper is based on a comparative literature review methodology, aiming to highlight the most efficient methods for automatic detection of fake news using Machine Learning and Deep Learning.

The study was conducted by identifying and selecting relevant scientific papers, published in specialized journals and conference in the field of artificial intelligence and natural language processing. This selection criteria included topic relevance, types od algorithms used, datasets used and evaluation metrics reported. The analysis covered both classical models, such as Logic Regresion, SVM, Random Forest, XGBoost, KNN, and advanced deep learning models, such as CNN, RNN, BERT. The models were compared according to the performance obtained in the analyzed papers.

4. Evaluation Metrics

Various metrics are used to evaluate the performance of algorithms, most of which are based on the confusion matrix. The confusion matrix is a table that shows and compares the actual values with the predicted outcomes of a classification model (Table 1).

Table 1. Confusion matrix

		Current values	
		1	0
Predicted values	1	True positive (TP)	False positive (FP)
	0	False negative (FN)	True negative (TN)

Accuracy measures the percentage of correctly classified instances out of the total number of instances analyzed. It is calculated as the ratio between the sum of True Positives and True Negatives and the

total number of instances. This indicator is particularly relevant in situations where the distribution of classes is balanced, when the number of examples in each category (e.g., fake news and real news) is approximately equal [17].

Precision indicates the proportion of instances correctly classified as positive, relative to the total number of instances predicted as positive. It is calculated as True Positives / (True Positives + False Positives). This indicator shows how accurate the positive predictions are and to what extent the model avoids misclassifying negative instances as positive [17].

The Recall (or True Positive Rate) measures the proportion of correctly identified positive instances in the total existing positive instances. It is calculated as True Positives / (True Positives + False Negatives). This indicator reflects the model's ability to detect all positive cases in the dataset, being especially useful when it is important to minimize omissions [17].

The F1 Score is the harmonic mean of precision and recall and gives a balance between precision and recall. This metric is particularly useful in unbalanced datasets, as it takes into account both the precision and the model's ability to detect all positive instances. In binary classification, True Positives (TP) is the number of events correctly classified as positive and False Positives (FP) is the number of events incorrectly classified as positive, True Negatives (TN) is the number of events correctly classified as negative and False Negatives (FN) is the number of events incorrectly classified as negative [17].

5. Machine Learning in detecting fake news

Machine learning is a branch of artificial intelligence which is concerned with the development of systems that can learn and adapt based on the data that they process. Algorithms are capable of generating predictions, recognizing patterns in data, and making decisions based on the information extracted. They are classified into supervised, unsupervised, and semi-supervised machine learning algorithms, with the first two categories being used most frequently [18].

Supervised learning algorithms are trained on labeled datasets, where the results are known, in order to be able to correctly estimate future events. They learn to correctly associate input data with expected outcomes so that they can make accurate predictions based on new datasets.

Logistic Regression is a binary classification algorithm that estimates the probability that an instance belongs to one of two classes (in this case, true or false). This model is often used for simple classification tasks, it predicts whether an item is true or false based on features extracted from the text. Logistic Regression gives efficient results when there is a linear relationship between the input features and the output label. Although this model is not extremely complex, it is often efficient for data with simple relationships and has been used as a benchmark for comparisons with other more advanced algorithms. The mathematical functions of the hypothesis of logistic regression and cost to obtain an optimal probability are represented as follows:

$$h_\theta(X) = \frac{1}{1+e^{-(\beta_0+\beta_1 X)}} \qquad [19]$$

$$Cost(h_\theta(x), y) = \begin{cases} \log(h_\theta(x)), & y = 1 \\ -\log(1 - h_\theta(x)), & y = 0 \end{cases}$$

Support Vector Machines (**SVM**) is a popular algorithm for classification problems, using a mathematical concept called *maximum edge*. It aims to find a hyperplane that separates the data in the two classes (true/false) by as large a margin as possible. SVM maximizes the

distance between the separating hyperplane and the nearest data points in each class (called support vectors). This large margin improves the generalization of the model. For cases where the data is not linearly separable, SVM uses a technique called kernel trick, which transforms the data into a higher dimensional space where linear separation becomes possible.

Random Forests are an ensemble of decision trees that help improve performance and reduce the risk of overtraining. Each tree in the forest is trained on a random subset of the dataset, obtained by the bootstrap method. At each node, a random subset of features is selected to determine the splitting criterion. The final prediction of the Random Forest model is made by majority vote (for classification) or by averaging the outputs (for regression) of all trees in the forest.

K-Nearest Neighbors (**KNN**) is an instance-based algorithm that classifies an article based on the nearest points in the dataset. KNN has been used in various studies to learn patterns of classification of fake news based on features extracted from the text. This is a simple and efficient algorithm, but it can be slower for large datasets because of the need to calculate distances between instances. Distances between two points can be calculated using the following formulas [19]:

$$Euclidean\ distance = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$$

$$Manhattan\ distance = \sum_{i=1}^{k}|x_i - y_i|$$

$$Minkowski\ distance = \left(\sum_{i=1}^{k}|x_i - y_i|^q\right)^{1/q}$$

AdaBoost (Adaptive Boosting) **and XGBoost** (Extreme Gradient Boosting) are two of the most popular boosting methods used to enhance the performance of classification and regression models. Boosting is a technique that combines several weak models (usually simple decision trees) to build a strong and robust model. The basic idea is that the models are trained sequentially, with each new model focusing on correcting errors made by previous models.

AdaBoost increases the weight of instances that have been misclassified by previous models, making them more influential in the next iteration. XGBoost, on the other hand, applies a more advanced, gradient-based boosting technique that trains decision trees in a highly efficient and parallelized manner. In a study [19], XGBoost had a significant impact on improving overall accuracy and performance due to its ability to manipulate complex data and reduce prediction errors. XGBoost was one of the algorithms that contributed to the outstanding results for fake news classification.

Decision Trees (**CART**) is an intuitive machine learning method that builds a tree-like model, where each internal node represents a feature-based question, each branch corresponds to a possible answer, and the leaves indicate a classification or a numerical value (in the case of regression). In classification tasks, the tree splits the data based on features to create subsets that are as "pure" as possible. Splitting decisions are typically based on measures such as entropy or the Gini index, which assess the quality of the partitions.

In the research [19], decision trees (CART) have been used as part of an ensemble, playing a significant role in creating accurate classifications, especially for categorical or textual data. Although decision trees are easy to interpret, they can be less effective in handling data with multiple meanings or in detecting complex relationships.

A study conducted by two researchers from India [20] compared the performance of Logistic Regression and SVM in

detecting fake news distributed on social platforms. The results of applying the SVM and LR models are:

Table 2. Results from applying SVM and LR models on a fake news dataset [20]

Model	**Accuracy**	**Precision**
SVM	0.91	0.89
LR	0.95	0.93

Model	***Recall***	***F1-Score***
SVM	0.73	0.75
LR	0.79	0.83

On a dedicated dataset, LR achieved 95.12% accuracy and 93.62% precision, while SVM had 91.68% accuracy and 89.20% precision, showing a slight superiority of LR. The authors proposed a novel framework called Novel Fake News Detection (NFND) based on Logistic Regression and emphasized the need to extend the datasets and integrate advanced technique such as POS tagging, word2vec, and topic modeling to improve the performance.

In another study, some researchers [5], created two distinct datasets: one generated by crowdsourcing, covering six diverse domains (sports, business, entertainment, politics, technology and education) and another collected from the web, focusing on celebrity news. The true news articles originated from trusted sources such as ABCNews, CNN and The New York Times, while the fake articles were written by trained workers on Amazon Machanical Turk to mimic journalistic style. For classification, a linear SVM was used with five-fold cross-validation, and the model performed exceptionally well using linguistic features such as readability, punctuation and the LIWC lexicon. The results showed high accuracy, sometimes even outperforming human annotators, although cross-domain generalization remained difficult, suggesting that structural and linguistic differences exist between fake news across different content areas.

On the other hand, Conroy and collaborators [6] proposed a complementary approach, integrating SVM into a hybrid system that combines linguistic technique with social network analysis. In this context, SVM vas applied to datasets containing both real and fake articles, including satirical and manipulated news, and demonstrated strong performance in identifying deceptive patterns. The authors concluded that this integration significantly improves fake news detection by leveraging both textual content and social context.

Both studies confirm the usefulness of SVM in automatic detection, as an efficient classification model based on linguistic features and as a part of complex hybrid system that combine different typed of data.

Acording to the study by Ahmad and others [19], the performance of several machine learning algorithms on four different datasets were evaluated. The obtained results were analyzed in order to compare the efficiency of each algorithm, depending on the characteristics of the datasets and their typology. The datasets used are open source and include both real and fake articles from various domains.

The first dataset, DS1 (ISOT Fake News Dataset), contains almost 45.000 articles, half are real, mainly from Reuters.com, and half are fake, sourced from disinformation websites, mostly political. The second dataset, DS2, available on Kaggle, includes more than 25.000 articles from various domains, split into training and test sets. The third dataset, DS3, also from Kaggle, has 3.300 articles from trusted sources such as CNN and The New York Times, as well as from untrusted sources focusing on sports, entertainment and politics. To allow for a more comprehensive evaluation, a combined dataset (DS4) was created, integrating all the articles from the previous sets.

Table 3 summarizes the average performance scores calculated across all four datasets to compare the overall effectiveness of the evaluated classification algorithms.

Figure 1 is a graphical representation of the average performance of the algorithms on all datasets [19].

Table 3. Average performance of learning algorithms [19]

Model	Average Precision	Average Recall	Average F1-Score
LR	0.93	0.92	0.92
LSVM	0.68	0.79	0.72
RF	0.80	0.80	0.79
KNN	0.70	0.67	0.68
AdaBoost	0.92	0.92	0.92
XGBoost	0.95	0.94	0.94
Decision Tree (CART)	0.94	0.94	0.94

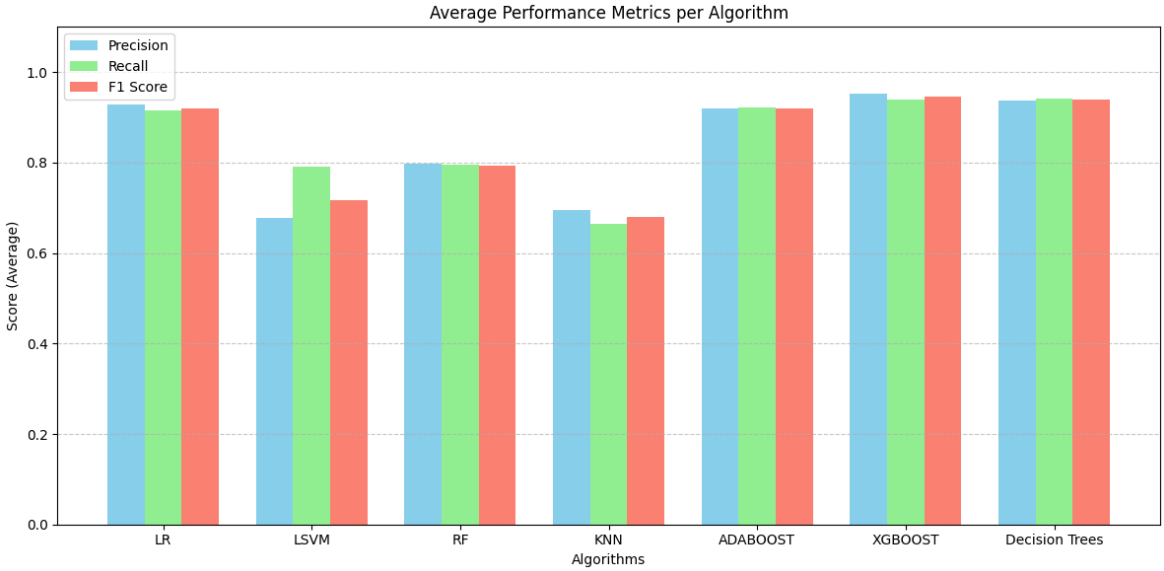

Fig. 1. Average performance of learning algorithms [19]

XGBoost ranked as the best performing algorithm, achieving an average F1 score of 0.95. Its superior performance is due to its ensemble boosting method and advanced regularization techniques that minimize classification errors systematically. AdaBoost and Decision Trees also demonstrated strong results, with F1 scores of 0.92 and 0.94 respectively, benefiting from adaptive learning mechanisms and hierarchical decision structures [19]. Logistic Regression maintained a constant performance (F1-score = 0.93), showing a particular efficiency on homogeneous datasets, while it showed a moderate decrease (to 0.87) on more heterogeneous data. This suggests its reliability for linearly separable problems, but also limitations in handling complex nonlinear relations. The analysis revealed suboptimal performance for linear SVM (F1 score = 0.73) and KNN (F1 score = 0.68). Linear SVM demonstrated a significant compromise between precision (0.68) and recall (0.79), while KNN performance was

affected by sensitivity to data noise and dimensionality [19].

A relevant example of the application of machine learning algorithms is presented in the study [21], published in the IOP Conference Series: Materials Science and Engineering. The authors used the LIAR-PLUS Master dataset, which contains veracity-labeled claims from the politifact.com platform. The dataset was preprocessed using the NLTK and SAFAR v2 libraries, applying operations such as text cleaning, tokenization, POS tagging and linguistic features extraction (e.g. average word length and adjective frequency). The study compares the performance of several classification algorithms, including Naïve Bayes, Random Forest, XGBoost, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree and Linear Regression.

The experimental results showed that XGBoost algorithm achieved the best performance, with an accuracy of 75.30%, an F1-score of 77%, a precision of 76% and a recall of 0.75%. It was followed by SVM, which obtained an accuracy of 73.20% and Random Forest with an accuracy if 72.50%.

Simpler models, like Naïve Bayes and KNN, performed worse, with accuracies of approximately 65% and 62%, while Linear Regression gave modest results, confirming its limitations when addressing complex text classification tasks.

Figure 2 is a graphical representation of the average results of all the algorithms [21]:

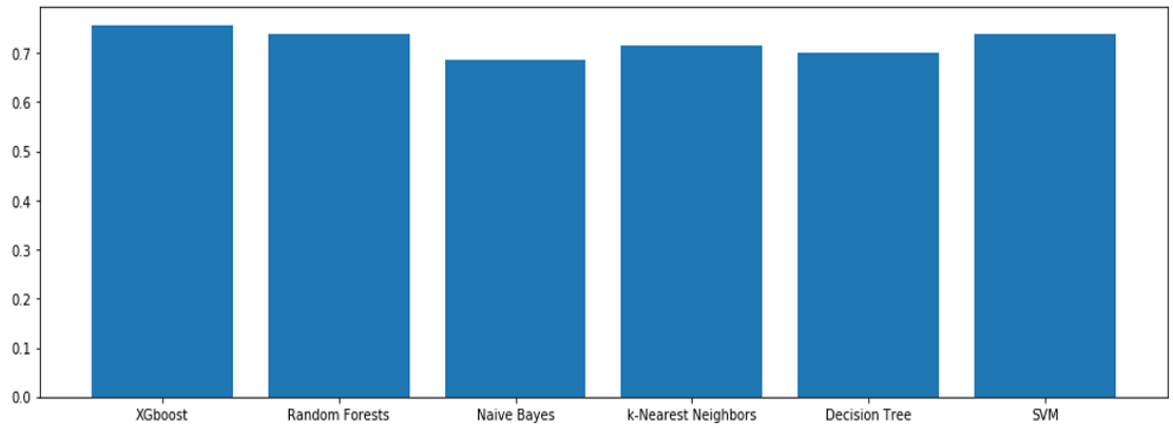

Fig.2. Accurancy Results of all the Algorithms

In addition to individual model testing, the authors also explored hybrid models by combining multiple classifiers through majority voting strategies. The results indicated an improvement in performance: the hybrid model composed of XGBoost, SVM and Random Forest achieved an accuracy of 81.20%, demonstrating increased robustness and reduced classification errors.

The comparative analysis underscored the importance of selecting an appropriate algorithm and complementary models can lead to superior performance in fake news detection tasks.

While supervised learning algorithms have shown promising results in the automatic detection of fake news, several studies have also highlighted their limitation, particularly regarding generalizability in novel contexts or across varied domains. For this reason, recent research has explored complementary methods that combine linguistic analysis and structural information.

Algorithms with unsupervised learning are trained on unlabeled datasets, where the correct results are not known. Their goal is to identify hidden patterns, structures or relationships in the data without knowing a previously provided output. These algorithms learn to organize and classify data on their own, discovering clusters or similar features, which can then be used to understand and interpret new data. This approach is useful when manual labeling is impossible. Although less precise than supervised methods, clustering provides valuable information in the absence of labels.

A key factor in one of the success factors behind machine learning algorithms is *parameter tuning*. This means that you will have to tune important parameters like the number of trees in a random forest, the level of regularization in a logistic regression or the number of neighbors in the KNN. The optimal choice of these parameters allows algorithms to become much more efficient in identifying false news, especially when using complex and large datasets.

Feature selection techniques, such as dimensionality reduction (e.g. PCA), can help improve results by eliminating irrelevant variables and preventing over-training. Appropriate adjustments of these parameters and features can lead to better performance, increasing the effectiveness of systems designed to detect fake news.

A relevant example of the application of unsupervised learning in fake news detection is presented in the study by Yang and colleagues [22], where they develop an original method based on probabilistic modelling. This research tackles the challenge of identifying fake news without using manually labeled datasets, which differentiates it from most existing approaches. The proposed model, called **UFD** (Unsupervised Fake News Detection), is based on the idea that users' interactions on social media, such as likes, retweets or comments on posts, may reflect their perceptions of a news story's veracity. They consider both news veracity and user credibility as latent variables and incorporate them into a generative model built on Bayesian networks. To validate the method's performance, they used two real datasets: LIAR and BuzzFeed News.

The UFD model demonstrated notably strong results. Compared to other unsupervised methods (Majority Voting, TruthFinder, LTM, CRH), the proposed model performed better.

For **Majority Vote** each news item gathers the opinions of verified users, and the most common version is considered true.

TruthFinder tries to find out which information is true, even if user opinions contradict each other. It analyzes the conflicts between opinions and calculates which information is most likely to be correct, even without knowing in advance what is true or false.

LTM is a model that recognizes that users can be wrong. It tries to discover the truth even if some people give wrong information, but it needs a simple structure in which each source says something about particular news item.

CRH assesses how correct each user is in general. If a user has given correct information many times, their opinion will count more in the final decision. This calculates how credible everyone is and determines which information is true.

On the LIAR set, UDF achieved an accuracy of 75.90%, while the other methods ranged between 58% and 64%. Additionally, the model attained an F1-score of 74.1%, exceeding the performance of the majority voting method of over 20 percentage points.

An important advantage of this approach is its dual capability: it not only estimates the veracity of news content but also assesses the credibility of users. This dual output enhances the model's reliability and its

capacity to extract valuable insights from unstructured data.

Machine Learning has shown considerable potential in detecting fake news, with choosing the right algorithms being the most important step in obtaining accurate results. Algorithms such as XGBoost, AdaBoost and logistic regression have proven effective in correctly classifying articles, while unsupervised learning can uncover hidden patterns in the data, bringing additional value to the detection process. Combining multiple algorithms can also help create a more robust model.

By integrating multiple algorithms and evaluating them across diverse datasets, it becomes possible to identify the most suitable models for the automatic detection, depending on the type of data, the complexity of the relationships between variables and the computational resources available. The reviewed studies confirm that ensemble models, such as XGBoost and Random Forest, offer an excellent balance between accuracy and robustness, like Logistic Regression, and can also yield competitive results in more constrained scenarios, with the added benefit of interpretability. Moreover, the adoption of hybrid models that combine linguistic features, contextual signals and meta-information extracted from social media represents a promising direction for enhancing the reliability of fake news detection systems.

However, the ultimate success of a system is highly dependent on continuous optimization of algorithms and hyperparameter tuning, which can make the difference between an accurate model and a less performing one. Future research should incorporate these aspects to increase the accuracy and adaptability of counter-disinformation systems.

6. Deep Learning in fake news detection

Another branch of Artificial Intelligence, related to Machine Learning, is *Deep Learning*, abbreviated DL. Deep Learning consists of learning methods that allow computers to learn independently, without the intervention of a human factor to define rules or knowledge. These models are structured as an artificial neural network because of their architecture which is made of interconnected nodes across multiple layers, similar to the neurons of a biological brain. The difference between DL and ML is that the latter involves the use of neural networks that have an input layer of neurons, an output layer and sometimes 1-2 hidden layers, whereas Deep Learning uses *deeper* neural networks, as the name suggests, with multiple hidden layers.

An example of a deep neural network can be seen in the figure below:

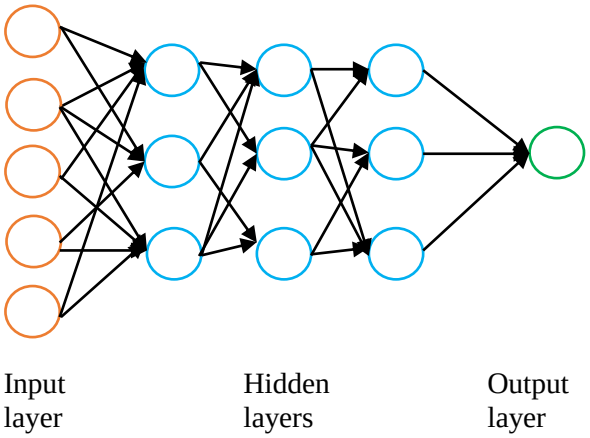

Fig. 3. Example neural network with multiple hidden layers

Natural Language Processing, or **NLP**, is another branch of AI, like ML and DL. Using NLP, computers can understand human language [23]. This technology is the basis of intelligent assistants. With NLP, these assistants can understand and reproduce the human language. NLP provides a number of data *pre-processing techniques* that are essential for ML and DL. These methods necessary especially for the analysis of a group of sentences, such as a news story. An NLP method is

tokenization in which a text is split into a sequence of tokens, consisting of words or parts of words. Another approach is the *Bag-of-Words* model, in which a data set is viewed as a multitude of words put together, ignoring their order. This calculates the frequency of occurrence of words in the text. One use of Bag-of-Words is searching for certain information on the Internet or other search engines. Bag-of-Words technique along with **TF-IDF** are vectorization methods often used in ML and DL models for fake news detection.

Another vector-based NLP technique is the **word embeddings** technique, which assigns different real number vectors to words, with the resulting representations being distributed, with a reduced dimension. This method is considered a more improved version of Bag-of-Words, the vector space being continuous and multidimensional. Models often used in fake news detection, pre-trained with word embeddings, are *GloVe* and *Word2Vec*. These have the ability to train very large datasets. Word2Vec represents performance in detecting syntactic relations between words, while GloVe does better with global semantic relations.

A first and common DL model is **CNN** (*Convolutional Neural Network*), first introduced by Kunihiko Fukushima [24]. Convolutional Neural Network implies the use of convolutional layers within the model. These layers filter the input data so that only relevant and useful information is extracted. CNN is basically used for image classification and recognition as input data. Convolutional layers may have other pooling layers attached to them. These layers reduce the size of images to decrease the number of parameters in a network. With other words, a pooling layer is a small portion of the input image and the convolutional operation helps extract its most important features.

A study by M. F. Mridha [25] highlights that *max pooling and average pooling* are the most commonly used functions in CNNs. Their research identified best parameters values for a CNN model in order to perform in detection of the fake news: the dense layer has 100 units, there are 100 filters, and the filter size is 5. *GlobalMaxPooling1D*, which performs global pooling, has the highest score among the CNN approaches, making it the most effective solution for detecting fake news.

As with any other learning model, certain problems arise, such as error reduction or *overfitting*, which occurs when model performance is poor on new data. One of the solutions to the overfitting problem is *regularization*, which is most common in detecting false news, according to Mridha's research.

Another often used model of Deep Learning is **RNN** (*Recurrent Neural Networks*), used for processing sequential data, such as speech and language. The structure of these neural networks includes recurrent links that *memorize* the information from previous execution of the same computation process. In this way, the final result depends on the results obtained during the whole process [24]. A problem encountered in the RNN application is the *disappearance of the gradient in time*.

A variation of the RNN model is RNN with Long Short Term Memory (**LSTM-RNN**) [17]. The added memory (LSTM) is intended to retain information from previous computations over a longer period. In detecting fake news it is important to understand the context, to look at the information as a whole for a correct classification, that is why the LSTM-RNN model is often used in this case. Nowadays, most of the fake news are appearing on social networks. In their study, Sahoo and Gupta [25] tried to detect such news items on the Facebook platform. They considered both the data of the posts and the account information of the people who published them, applying the LSTM-

RNN model. However, due to too much data, the runtime was too long.

Another variation of the RNN model is **GRU** (*Gated Reccurent Unit*), similar to LSTM, but with a simpler architecture. GRU needs fewer parameters, making the train process more efficient. The main difference between GRU and LSTM model is how they manage the memory. LSTM has a separated memory part, which is updated through three gates (input, output and forget). On the other hand, GRU combines the memory and the hidden parts into one, using only two gates (reset and update). Through these gates, GRU model chooses what information is modified. At a conference from 2024, Elfaik and Nfaoui [26] did a comparative study between GRU and other Deep Learning methods, LSTM and RNN, used to detect fake news. They used ISOT Fake News Dataset and after cleaning the data and applying selected methods to identify the fake news, they observed that GRU model had a higher accuracy compared to the others, as we can observe in Table 4 bellow. This means that GRU model is highly effective for detecting fake news with minimal manual features extraction.

Table 4. GRU vs. LSTM and RNN – accuracy values [26]

Model	**Accuracy**
LSTM	0.9969
RNN	0.7448
GRU	0.9983

The LSTM of a recurrent neural network can also be bidirectional, related to the **BiLSTM-RNN** model. This is a variant of the traditional RNN model, but which provides a memory that allows data to be processed from the beginning to the end and from the end to the beginning. BiLSTM-RNN is used when both the past and the future need to be analyzed, it is more powerful than the classical LSTM, but also more expensive.

One pre-trained DL model is the **BERT** model, short for Bidirectional Encoder Representation from Transformers. This model is based on transformers, as the name suggests, and was introduced by Google in 2018 [17]. Transformers have the ability to process a word by taking into account its relationships with other words in the text, allowing the model to understand its context. Thus, BERT is used for natural language processing (**NLP**) in many situations, like translating languages.

As with RNN, BERT has several variations, one of which is **ALBERT**, which is derived from the *A Lite BERT model*, considered to be efficient to use for false news detection. In contrast to BERT, ALBERT uses a smaller projection layer and applies the *Weight Sharing principle*, which divides the weights among all layers, thus reducing the number of parameters.

RoBERTa (Robustly Optimized BERT Approach) is another improved variant of the BERT model, introduced in 2019 by Facebook AI researchers. The difference between the two models is that RoBERTa can process a much larger volume of data and uses improved training procedures. This makes RoBERTa a more powerful tool to use, especially for fake news detection. However, in 2020, following a study, the **FakeBERT** model specifically designed for fake news detection was introduced. FakeBERT is based on the architecture of the BERT model, and the difference costs in the training set containing fake and real news [27]. Another specific feature of FakeBERT is the use of the back-translation technique. This technique involves translating a real news text into a language, and then translating it back into the original language, generating synthetic data that is added to the training set.

7. Deep Learning vs. Machine Learning in fake news detection

Machine Learning and Deep Learning methods show performance in identifying fake news, as we can observe in this paper. ML utilizes classical algorithms such as SVM or Random Forest, while DL uses the principle of deep neural networks such as CNN and RNN. In contrast to ML, DL models have the ability to process large datasets containing raw texts, without human intervention, learning automatically. However, due to its higher performance, Deep Learning involves higher costs in terms of resources and processing time. In contrast, traditional models can be trained and deployed much faster on modest hardware, making them a more practical solution in low-resource environments.

Both branches of Artificial Intelligence have performed well in fake news detection, but their performance differs depending on the details of the problem. For example, if one wants to classify a text as fake or not, then an ML model can be used, but if one wants to analyze a large dataset, such as all the news posted on a web page, then an DL model is better to use, due to its improved ability to work on large datasets.

Several studies have been conducted over time comparing different ML and DL models applied on the same datasets. The Deep Learning methods presented earlier in this paper are the most common methods used in different studies on fake news detection. An example of such work is the comparative study of Deep Learning and Machine Learning methods applied to identify fake news, published in the *Carol* National Defense University Bulletin [17]. The datasets underlying the research include fake news from ISOT, BuzzFeed and PolitiFact websites. DL models, also described in this paper, were applied and results obtained as follows:

Table 5. Results from applying DL models on a fake news dataset [17]

Model	***Accuracy***	***F1 Score***
RoBERTa	0.99	0.99
LSTM-RNN	0.96	0.97
BiLSTM-RNN	0.98	0.97
ALBERT	0.97	0.97
FakeBERT	0.98	0.99
BERT	0.98	0.98
CNN	0.96	0.96

In the table above, we can observe that after applying the models, the best performing model in identifying fake news is RoBERTa, with accuracy and F1 score values very close to 1 and 0.99, respectively. The difference between RoBERTa and the other models is small, which shows that all models were able to correctly identify almost all fake news, the worst performing model being CNN, with an accuracy and F1 score of 0.96.

A case study is the work of S. Repede [17], mentioned earlier in the paper. He applied on 4 datasets several machine learning models. Analyzing the evaluation metrics, it was found that the most efficient model, with the highest performance, is the *RoBERTa* model, with values above 0.99, very close to the maximum value 1.

Another research conducted by Arshad Ali and Maryam Gulzar [28] was based on the detection of fake news on social networks related to the COVID-19 pandemic. In this research, the authors tried to combine two machine learning models, **BERT** (DL) and **SVM** (ML), along with an evolutionary algorithm, **NGSA-II** (Non-dominated Sorting Genetic Algorithm II), to obtain better results, thus creating a new hybrid model. After preprocessing the COVID-19 news dataset, different traditional ML and DL models were applied, including:

Table 6. Results applying classical ML and DL models on the COVID-19 dataset [28]

Model	*Accuracy*	*Precision*
SVM	0.60	0.65
Random Forest	0.68	0.63
K-Neighbour	0.64	0.69
BERT	0.63	0.72
CNN	0.72	0.73
Model	***Recall***	***F1 Score***
SVM	0.71	0.55
Random Forest	0.7	0.66
K-Neighbour	0.65	0.7
BERT	0.63	0.67
CNN	0.82	0.77

We can observe that the applied models have an average performance, the best being CNN, with the highest evaluation metrics values. After the application of the proposed hybrid model, BERT+NSGA-II+SVM, it was found that it performs much better than the traditional models applied individually. The F1 score (0.83) and accuracy (0.8) have values above 0.8, recall has a value of 0.9, and precision is 0.76. All metrics have higher values than those of the CNN model, the best performing of the traditional models applied.

At a conference, Chang [29] conducted a study on different Machine Learning and Deep Learning algorithms used to combat fake news. He used the ISOT dataset containing fake and real news. After pre-processing it, the author compared 15 algorithms, and the results showed that Deep Learning models performed the best, especially the BERT model with 99.95% accuracy. In second and third place were the BiLSTM and LSTM models, showing the superiority of Deep Learning on this dataset. Regarding Machine Learning, the best result belongs to the SVM model, with an accuracy of 98.65%.

A comparative study was carried out by Kishwar and Zafar [30] based on a dataset of news about Pakistan. The two researchers used Google Fact Checker along with a series of words to extract relevant news. In addition to Google Fact Checker, data was also extracted from other sources such as Kaggle. After applying different Machine Learning models and technologies, Deep Learning was found to perform better in identifying fake news. The CNN model in conjunction with GloVe achieved an F1 score of 0.93, while LSTM scored 0.94. An interesting thing introduced in the study is a questionnaire completed by 57 people who had to categorize 10 news stories as fake or real. The results of the questionnaire showed that most of the fake news stories correctly identified by those individuals were also classified as fake by the models applied. However, there were many misclassifications of real news stories as fake because they had a similar writing style to fake news, with information that appeared to be false. Thus, it was shown that ML models can overcome human judgment.

Similar to the previously mentioned studies, Alghamdi, Lin and Luo [31] did a study in which, on several fake news

datasets, LIAR, PolitiFact, GossipCop and COVID-19, they applied different models of both ML and DL. LIAR is a rather voluminous dataset used for identifying fake news, composed of US political statements. Similarly, PolitiFact is a dataset of political statements available on the website of with the same name. On the other hand, GossipCop contains data on celebrity news stories, and COVID-19, as the name suggests, contains news about the COVID-19 pandemic, both sets categorized as real or fake. Their study becomes even more interesting because they applied the same model several times, each time using different pre-processing techniques. Their results showed that the classical ML models, applied together with TF-IDF, perform best on the LIAR dataset, compared to more advanced models such as DL or hybrid models. In contrast, on the PolitiFact dataset, RoBERTabase performs best with a high F1 score of 0.93. Regarding the GossipCop dataset, classical ML models performed better, while on the COVID-19 dataset the BERTbase model, belonging to Deep Learning, performed best. This study shows that there is no universal ideal method that gives the same performance on all datasets, as context as well as other factors (resources, data volume and so on) matter a lot.

8. Conclusion

The problem of the spread of fake news today is growing. People spend much of their time on social media platforms like Facebook, Instagram and TikTok. For various reasons, certain individuals or *bots* post various news, ads or videos containing false information with the aim of mass manipulation or monetization. Therefore, the use of a mechanism to identify false information is essential. Both Machine Learning models, such as Random Forest and SVM, and Deep Learning models, such as RNN and FakeBERT, can be used for this purpose, as we have seen in this paper.

Machine Learning and Deep Learning perform well on this false news topic. However, depending on the size of the dataset, the complexity of the situation and the available resources, an ML model may be more efficient than a DL model or vice versa. It has been shown that for large datasets, in particular fake news datasets, Deep Learning models perform better due to their LSTM, BiLSTM and BERT methods. However, technologies are evolving, and hybrid solutions that have appeared in recent years, such as the BERT+NSGA-II+SVM model proposed and analyzed by Ali and Gulzar [28], which combines ML models with DL models, NLP techniques and other technologies, perform much better and represent the future of combating online misinformation.

References

[1] K. S. A. W. S. T. J. &. L. H. Shu, „Fake News Detection on Social Media: A Data Mining Perspective," *ACM SIGKDD Explorations Newsletter,* vol. 19, nr. 1, pp. 22-36, 2017.

[2] U. A. f. C. ENSIA, "ENISA THREAT LANDSCAPE 2022," *European Union Agency for Cybersecurity,* 2022.

[3] D. P. Supervisor, „Fake News Detection," 2023. [Interactiv]. Available: https://www.edps.europa.eu/press-publications/publications/techsonar/fake-news-detection_en. [Accesat 27 3 2025].

[4] Y. C. N. J. C. V. L. Rubin, "Deception detection for news: three types of fakes," *Proceedings of the Association for Information Science and Technology,* vol. 52, no. 1, pp. 1-4, 2015.

[5] K. A. L. R. M. V. Perez-Rosas, "Automatic Detection of Fake News,"

arXiv preprint arXiv:1708.07104, 2017.

[6] V. L. R. Y. C. N. J. Conroy, "Automatic deception detection: Methods for finding fake news," *Proceedings of the Association for Information Science and Technology*, vol. 52, no. 1, pp. 1-4, 2015.

[7] M. S. A. M. F. SM H. Dadgar, "A novel text mining approach based on TF-IDF and Support Vector Machine for news classification," *Proceedings of IEEE International Conference on Engineering and Technology* , pp. 112-116, 2016.

[8] S. S. Y. L. N. Ruchansky, "CSI: A hybrid deep model for fake news detection," *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 797-806, 2017.

[9] S. W. H. L. K. Shu, "Understanding User Profiles on Social Media for Fake News Detection," *IEEE Conference on Multimedia Information Processing and Retrieval*, 2018.

[10] J. C. Y. Z. J. L. Z. Jin, "News Verification by Exploiting Conflicting Social Viewpoints in Microblogs," *Thirtieth AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, pp. 2972-2978, 2016.

[11] Q. L. R. X. M. L. C. R. H. Y. Long, "Fake news detection through multi-perspective speaker profiles," *Proceedings of the Eighth International Joint Conference on Natural Language Processing*, pp. 252-256, 2017.

[12] M. R. C. Janze, "Automatic detection of fake news on social media platforms," *Proceedings of 21st Pacific-Asia Conference on Information Systems*, 2017.

[13] J. G. C. Buntain, "Automatically identifying fake news in popular twitter threads," *Proceedings of the IEEE International Conference on Smart Cloud, New York*, pp. 208-215, 2017.

[14] W. H. C. J. F. G. L. H. Zhu H., "Information dissemination model for social media with constant updates," *Physica A: Statistical Mechanics and its Applications*, vol. 502, pp. 469-482, 2018.

[15] S. M. G. R. A. M. A. K. S. Tschiatschek, "Fake news detection in social networks via crowd signals," *Proceedings of the World Wide Web Conferences, France*, pp. 517-524, 2018.

[16] S. A. E. E. P. G. B. Guacho, "Semi-supervised content-based fake news detection using tensor embeddings and label propagation," *Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 322-325, 2018.

[17] R. B. Ș. E. Repede, "O comparație a modelelor de inteligență artificială folosite pentru detectarea știrilor false," *Buletinul Universităţii Naţionale de Apărare »Carol I«,* no. 1, pp. 81-99, 2023.

[18] Oracle, „Ce este machine learning?," 2025. [Interactiv]. Available: https://www.oracle.com/ro/artificial-intelligence/machine-learning/what-is-machine-learning/. [Accesat 27 Martie 2025].

[19] M. Y. S. Y. M. O. A. I. Ahmad, "Fake News Detection Using Machine Learning Ensemble Methods," *Complexity*, vol. 2020, no. 1, 2020.

[20] M. A. N. Leela Siva Rama Krishna, "Fake News Detection System using Logistic Regression and Compare Textual Property with Support Vector Machine Algorithm," *Proceedings of the International Conference on Sustainable Computing and Data Communication Systems* , 2022.

[21] N. A. H. S. M. R. Z Khanam, "Fake News Detection Using Machine Learning Approaches," *IOP Conference Series: Materials Science and Engineering*, 2021.

[22] S. S. K. W. S. G. R. W. F. &. L. H. Yang, "Unsupervised Fake News Detection on Social Media: A Generative Approach," *Proceedings of the AAAI Conference on Artificial Intelligence,* vol. 33, no. 1, pp. 5644-5651, 2019.

[23] Eppright, „Ce este NLP (procesarea limbajului natural)?," 2021. [Interactiv]. Available: https://www.oracle.com/ro/artificial-intelligence/what-is-natural-language-processing/.

[24] O. Y. A. U. a. Y. D. M. Coskun, "An Overview of Popular Deep Learning Methods," *European Journal of Technic, Vol. 7, Number 2,* pp. 165-176, 2017.

[25] J. K. M. A. H. M. M. M. a. M. S. R. M. F. Mridha, "A Comprehensive Review on Fake News," *IEEE Access,* pp. 156151-156170, 2021.

[26] H. N. H. Elfaik, "Automatic Detection of Fake News Using Gated Recurrent Unit Deep Model," in *5th International Conference on Innovative Data Communication Technologies and Application*, 2024.

[27] G. a. P. N. R. K. Kaliyar, „FakeBERT: Fake news detection in social media with a BERT-based deep learning approach," *Multimedia Tools and Applications,* pp. 11766-11788, 2021.

[28] M. G. A. Ali, "An Improved FakeBERT for Fake News Detection," *Applied Computer Systems, Vol. 28, No. 2,* pp. 180-188, 2023.

[29] L. Chang, "Comparison of Machine Learning and Deep Learning Algorithms in Detecting Fake News," in *Proceedings of the 28th World Multi-Conference on Systemics, Cybernetics and Informatics*, 2024.

[30] Z. A. Kishwar, "FakenewsdetectiononPakistaninewsusingmachinelearninganddeep learning," *Expert Systems With Applications,* vol. 211, pp. 1-10, 2023.

[31] Y. L. S. L. J. Alghamdi, "A Comparative Study of Machine Learning and Deep Learning Techniques for Fake News Detection," *Information,* pp. 1-28, 2022.

Gabriela Chiriac graduated in 2020 from the Faculty of Accounting and Management Information Systems and she is currently a final year student in the master's program Databases – Business Support at the Faculty of Cybernetics, Statistics and Economic Studies, Bucharest University of Economic Studies. Professionally, she works as a Data Engineer, focusing on developing and maintaining ETL flows, managing databases and providing technical support for internal stakeholders.

Ada Maria Catina graduated in 2020 from the Faculty of Cybernetic, Statistics and Economic Studies, specializing in Economic Informatics. She is currently in her final year of the master's program Databases – Business Support at the Faculty of Cybernetics, Statistics and Economic Studies, Bucharest University of Economic Studies. Her academic interests focus on databases, data science and the application of machine learning techniques.