Fake News Detection Using Passive Aggressive Classifier (PAC) & Neural Network: A Hybrid Approach

Om Prakash Mohanta
Computer Science and Engineering
C.V. Raman Global University
Bhubaneswar, India
pikunmohanta50@gmail.com

Gourav Swain
Computer Science and Engineering
C.V. Raman Global University
Bhubaneswar, India
gouravswain03@gmail.com

Kumar Ashutosh
Computer Science and Engineering
C.V. Raman Global University
Bhubaneswar, India
kumarashutosh7917@gmail.com

Arpit Padhy
Computer Science and Engineering
C.V. Raman Global University
Bhubaneswar, India
arpit.padhy789@gmail.com

Sampa Sahoo
Computer Science and Engineering
C.V. Raman Global University
Bhubaneswar, India
sampaa2004@gmail.com

***Abstract*—In this digital age, the proliferation of fake news represents a significant challenge. Our research introduces a hybrid detection system that employs an optimized Passive Aggressive Classifier (PAC), fine-tuned through various parameters, alongside Artificial Neural Networks (ANN). This enhanced PAC is designed to identify and counteract misinformation on social media and anonymous websites. By analyzing articles verified by fact-checkers, the system accurately pinpoints linguistic signs of fake news, ensuring precise content classification. Our initiative supports informed public dialogue by providing immediate detection and rationale for content identified as potentially false. Trained on a comprehensive dataset that includes text, titles, and authorship, our robust framework adeptly evaluates the veracity of online articles. Through the application of advanced machine learning and natural language processing techniques, we tackle the rapid spread of false information, aiming to protect the digital information sphere and guarantee access to dependable and accurate news.**

***Index Terms*—Fake News Detection, Passive Aggressive Classifier, Neural Networks, Natural Language Processing, Machine Learning**

I. Introduction

In an age dominated by the omnipresence of social media platforms and instant messaging services, distinguishing veracity from falsehood has become an increasingly daunting task. Despite the indispensable role of traditional fact-checking methodologies, they often encounter formidable obstacles such as time constraints and financial limitations [1], which impede their efficacy in combating the rampant dissemination of misinformation. This conundrum starkly underscores the urgent imperative for pioneering solutions to fortify the integrity of information in the digital epoch.

To confront this multifaceted challenge effectively, it is imperative to embark upon a comprehensive and decisive approach. Such an approach necessitates the collaborative efforts of stakeholders from diverse sectors, including technology, media, academia, and government [2]. Innovative tools and platforms can be developed to empower individuals with the capability to discern fact from fiction amidst the deluge of information inundating cyberspace. It requires the collective efforts of stakeholders from various sectors, including technology, media, academia, and government. In response to the pressing need for a dependable means of distinguishing authentic from false news, our solution harnesses technologies such as Passive Aggressive Classifiers (PAC), and Artificial Neural Networks (ANN). Integrated with the Rapid API, our methodology begins by identifying relevant article URLs and employs advanced feature extraction techniques using PAC and ANN combined. This innovative framework offers a rapid and robust approach to combating misinformation, crucial for preserving information integrity in the digital realm. By leveraging state-of-the-art tools and techniques, our solution aims to address the escalating challenge of fake news, empowering users to navigate the information landscape with confidence and accuracy.

The PAC algorithm's meticulous content classification achieves an impressive 98% accuracy, boosting user confidence and streamlining news article validation. Integration of advanced machine learning and deep learning models enables users to submit links for comprehensive analysis, empowering informed decision-making to combat digital misinformation effectively [3]. This comprehensive strategy not only bolsters online information credibility but also nurtures a more informed and resilient society amidst widespread misinformation. By leveraging cutting-edge technologies and promoting critical thinking, our approach addresses the urgent need for discerning truth in the digital age, fostering a culture of accountability and reliability in information dissemination.

II. Leveraging Passive Aggressive Classifier and Artificial Neural Networks for Fake News Detection

Artificial neural networks (ANNs) are computational models inspired by the structure and functionality of biological neural networks. They consist of interconnected nodes, or neurons, organized into layers, with each neuron performing simple computations. ANNs are trained using algorithms that adjust the weights and biases of connections between neurons to minimize the difference between predicted and actual outputs.

Passive Aggressive (PA) classifiers belong to a family of online learning algorithms, adept at handling classification tasks where data arrives in chunks or streams [4]. Unlike traditional algorithms, PA classifiers do not maintain a constant learning rate; instead, they adjust based on the accuracy of the classification.

The PA algorithm's unique name derives from its behaviour when it classifies a sample correctly, it remains 'passive,' making minimal changes. However, upon misclassification, it turns 'aggressive,' updating its parameters to correct the mistake. This adaptive nature makes PA particularly suitable for dynamic environments where data distributions might change over time, such as news streams.

Originally devised for text categorization tasks, the PA algorithm has gained traction in fake news detection due to its ability to handle large datasets efficiently [5]. By iteratively processing data and updating model parameters based on the most recent examples, PA ensures adaptability and accuracy, essential for discerning subtle nuances between genuine and fabricated news content.

A. Mathematical Model Of Passive-Aggressive Classifier

The foundational concept behind the PA classifier is its adaptive learning mechanism. When a news article's features are fed into the model, it classifies the content based on the current model parameters. If correct, the model remains passive; if incorrect, it adjusts its parameters aggressively to rectify the error. Mathematically, the update rule for the PA algorithm can be expressed as:

$$\omega = \omega + \alpha * (y - \hat{y}) * x \tag{1}$$

ω represents the weight vector.
α is the learning rate, which can adapt based on the model's performance.
y is the true label, and $\hat{y}$ is the predicted label.
x is the feature vector of the news article.

For fake news detection, this process continues iteratively, ensuring that the model fine-tunes itself based on incoming data. The classifier's efficiency lies in its ability to swiftly adapt to changing patterns, making it a robust tool in the fight against misinformation.

B. Mathematical Model Of Artificial Neural Network

A mathematical model of an artificial neural network (ANN) is crucial for understanding its behavior, training process, and predictive capabilities. ANNs are computational models inspired by biological neural networks in the human brain. They consist of interconnected nodes, called neurons, organized into layers. The mathematical model of an ANN typically involves defining the architecture, activation functions, weights, biases, and the learning algorithm.

$$y_i^l = f\left(\sum_j w_{ij}^l \cdot y_j^{l-1} + b_i^l\right) \tag{2}$$

y_i^l is the output of neuron i in layer l.
f is the activation function.
w_{ij}^l is the weight of the connection between neuron
i in layer l, and neuron j in layer $l-1$.
y_j^{l-1} is the output of neuron j in the previous layer.
b_i^l is the bias term for neuron i in layer l.

C. Literature Study

Vitaly Klyuev and his colleagues conducted a comprehensive examination of various approaches for detecting fake news, focusing on semantic analysis through natural language processing (NLP) and text mining methodologies. Additionally, they explored strategies for automated verification and identification of bots within social media platforms [6]. In contrast, Oshikawa and collaborators presented a survey exclusively centred on fake news detection methodologies, with a specific emphasis on scrutinizing NLP-based techniques [7]. Meanwhile, Collins and colleagues delved into the diverse manifestations of fake news, exploring a range of its iterations. Additionally, they examined contemporary strategies aimed at curbing the dissemination of fake news across social media platforms [8]. Mariana Caravanti de Souza's PU-LP Algorithm employs a limited set of labeled fake news to classify content, achieving state-of-the-art results through a network-based approach and semi-supervised learning [9].

TABLE I
DATASET COMPARISON TABLE

Name Of Dataset	Domain	Concept	No of class
Fake_or_real_news [10]	Politics, Society	Text	2
LIAR [11]	Politics	Text	6
BREAKING! [12]	Society,Politics	Text,Image	2,3
Stanford Fake News [13]	Society	Text,Image	2
BuzzFace [14]	Politics,Society	Text	4
FacebookHoax [2]	Science	Text	2
Fact-Checking [15]	Politics,Society	Text	5
ISOT [16]	Business,Crime	Text	2
FakeHealth [17]	Health	Text	2

Iftikhar Ahmad's paper underscores the importance of detecting counterfeit news due to its extensive impacts on sectors like politics and financial markets. Additionally, it highlights how ensemble techniques can enhance the accuracy of fake news categorization. This collection of research reflects a multifaceted exploration of fake news detection methods [18]. In TABLE I, various datasets used by researchers is presented. The datasets were compared based on concept type, domain and number of classes.

ISOT: Comparing Reuters' real news with fake news from sites marked by PolitiFact and Wikipedia [16].

LIAR: A collection of 12,836 political statements from online streaming and social media platforms (Twitter and Facebook) between 2007 and 2016 [11].

BREAKING!:A dataset derived from the Stanford Fake News dataset and the BS detector dataset, containing web-based news about the 2016 US presidential election [12].

Stanford Fake News: A dataset of fake news and satire stories, featuring exaggerated praise or criticism of a person, conspiracy theories, racist themes, and attacks on credible sources [13].

BuzzFace A dataset of 2263 news articles and 1.6 million comments on Facebook about social and political issues from July 2016 to December 2016. This dataset was updated in September 2016 [14].

FacebookHoax: A dataset of 15,500 science-related hoaxes on Facebook from July 2016 to December 2016. This dataset also tracks posts with more than 2.3 million likes [2].

III. IMPLEMENTATION DETAILS

In relation to the workflow illustrated in Fig. 1, we provide an overview of our proposed framework, followed by detailed descriptions of algorithms utilized, datasets employed, and the metrics employed for performance evaluation.

A. About Dataset

This study leveraged a dataset sourced from Kaggle [10], encompassing 20,800 articles categorized as either fake or real news. The dataset we have gathered includes critical attributes such as textual content, title, and author's name, which serve as foundational elements for subsequent analysis and model training. Data Exploration and Preprocessing: Initial dataset exploration revealed anomalies, such as missing values, addressed through meticulous data-cleaning procedures. The dataset exhibited an equitable distribution, with 50% fake news and 50% genuine articles, enhancing its reliability. Model Evaluation Dataset Split: The dataset was systematically split into training (80%) and testing (20%) sets, ensuring a robust evaluation of machine learning models. Stratified sampling methods maintained a consistent proportion of fake and real news articles across both subsets.

B. Data Loading and Preprocessing

The dataset from Kaggle comprises 20,800 articles labeled as fake or real news, featuring text, title, and author attributes. This dataset includes attributes such as text, title, and author information. Addressing missing values and duplicates, we ensured data integrity through thorough cleaning and NLP techniques like tokenization and stop-word removal [19]. Feature engineering extracted key elements like linguistic and social context, aiding in categorizing credibility. Using stratified sampling, the dataset is divided into 80% training and 20% testing sets for balance model evaluation.

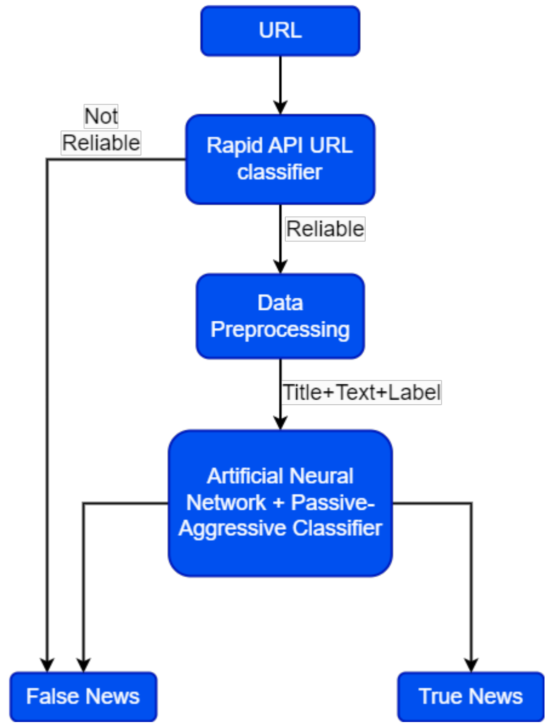

Fig. 1. Workflow Model

C. Text Tokenization and Padding

We used the NLTK library for text tokenization, converting raw data into unigrams to efficiently capture word frequencies [20]. To handle varied text lengths, we applied zero padding, standardizing sequences to a 300-word maximum. This approach balanced preserving article content while ensuring uniformity for machine learning model compatibility.

D. Converting Text to Vectors

In the context of the Passive Aggressive algorithm, the conversion of text into numerical vectors is a crucial preprocessing step for classification tasks. This process is facilitated by the utilization of CountVectorizer, a technique employed by the Passive Aggressive Classifier (PAC). CountVectorizer transforms textual documents into numerical representations by tabulating the frequency of occurrence of each word. These numerical vectors serve as the input for the algorithm, enabling it to discern patterns and make informed predictions based on the textual features. By employing this method, the PAC algorithm efficiently analyzes text data, contributing to its effectiveness in various applications such as fake news detection and sentiment analysis.

E. PAC Model Creation

The Passive-Aggressive (PA) Classifier is an online learning method used for binary classification tasks. Unlike traditional models, it doesn't store past data. Instead, it adapts dynamically based on prediction outcomes. When predictions are correct, it remains passive; when errors occur, it becomes aggressive and adjusts model parameters. Key hyperparameters include the regularization parameter, which controls aggressiveness, and the maximum number of iterations. PA is efficient for real-time applications and handles evolving data patterns effectively [21].

F. ANN Model Creation

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, 773, 100)          50322300

 bidirectional (Bidirectiona  (None, 773, 128)         84480
 l)

 dropout (Dropout)           (None, 773, 128)          0

 bidirectional_1 (Bidirectio  (None, 64)               41216
 nal)

 dropout_1 (Dropout)         (None, 64)                0

 dense (Dense)               (None, 64)                4160

 dropout_2 (Dropout)         (None, 64)                0

 dense_1 (Dense)             (None, 1)                 65

=================================================================
Total params: 50,452,221
Trainable params: 50,452,221
Non-trainable params: 0
```

Fig. 2. Architecture of the Sequential ANN

Artificial Neural Networks (ANNs) simulate the human brain, performing tasks like image recognition, natural language processing, and pattern detection. These networks consist of interconnected nodes (artificial neurons) in input, hidden, and output layers. Each node processes input using weights, biases, and activation functions. Our project implemented a sequential ANN to process text inputs through an 8-dimensional matrix. It included bidirectional layers to capture dependencies in both directions. Dropout layers prevented overfitting. Two dense layers with Rectified Linear Unit (ReLU) activations aided pattern recognition, and the final output layer used a sigmoid activation function. Optimized with the Adam optimizer and binary cross-entropy loss, the model had over fifty thousand trainable parameters. [5]. In Fig. 2 provides a visual representation of this ANN architecture, illustrating the various layers and their connections. The embedding layer, bidirectional layers, and dense layers are clearly depicted, emphasizing the flow of information through the network. .

G. Fake News Detection & User Interaction via Flask Framework

Developed a Flask web application that combines advanced machine learning methods with a user-friendly interface for Fake News Detection. Utilizing the Passive Aggressive Classifier (PAC) and CountVectorizer, users can enter news article URLs to check their validity. The application is enhanced by HTML, CSS, and libraries such as Flask, nltk, and scikit-learn, offering a visually appealing interface and smooth functionality. An integral feature of this tool is its interactive dashboard, which presents a comprehensive figure of the user interface, referred to as Fig. 3. This figure illustrates the process flow from URL entry to validity assessment, including visual cues for classification results and statistical insights into the data analyzed. Demonstrating our commitment to fighting digital misinformation, we provide users with a reliable and straightforward solution. In our proposed framework, depicted

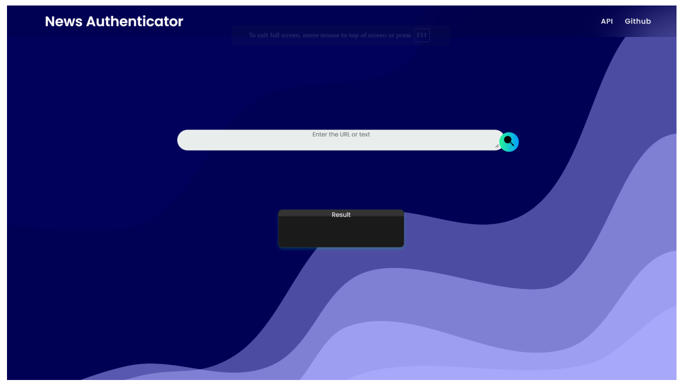

Fig. 3. User Interface

in Fig. 1, we aim to advance existing research by integrating ensemble techniques alongside diverse linguistic feature sets [22].This innovative approach is tailored for the classification of news articles originating from diverse domains, distinguishing between genuine and fabricated content.

H. Implementation of Rapid API

Rapid API presents developers with a versatile platform that facilitates the seamless integration of various APIs, spanning cybersecurity and threat intelligence services. Although not specifically tailored to fake website detection, its suite of tools offers invaluable insights into the reputation and vulnerabilities of websites [23]. By integrating with Rapid API, our system gains access to real-time data and analytics, empowering users to make informed assessments of website legitimacy. This integration not only enhances online security measures but also equips users with the necessary tools to navigate the digital landscape with confidence, mitigating risks associated with fraudulent websites and bolstering overall cyber defense strategies.

I. Output

The output of this process is the classification of each news article: Fake News: If the combined prediction leans toward fake. Real News: If the combined prediction leans toward real. In summary, by checking URL reliability, preprocessing data, and leveraging both ANN and PAC models, we aim to accurately classify news articles as fake or real. The combination of these models enhances the overall prediction performance.

IV. Results

Assessing the effectiveness of algorithms involved the utilization of various metrics, primarily derived from the confusion matrix. This matrix serves as a structured representation of a classification model's performance on the test dataset, encapsulating four key parameters: true positives, false positives, true negatives, and false negatives.

A. Accuracy

Accuracy is a performance metric used to evaluate the effectiveness of a classification model. It measures the proportion of correctly classified instances out of all instances in the dataset.

B. Precision

Precision is a performance metric used in binary classification tasks that measures the proportion of true positive predictions among all positive predictions made by the model. It indicates the model's ability to avoid false positives.

C. Recall

Recall signifies the entirety of positive classifications among the true class. In our context, it denotes the count of articles forecasted as true relative to the overall count of true articles.

D. F1 Score

The F1 Score encapsulates the balance between precision and recall, computing the harmonic mean of both metrics. By considering both false positive and false negative observations, it provides a comprehensive assessment. The F1 Score can be determined using the following formula:

E. Findings

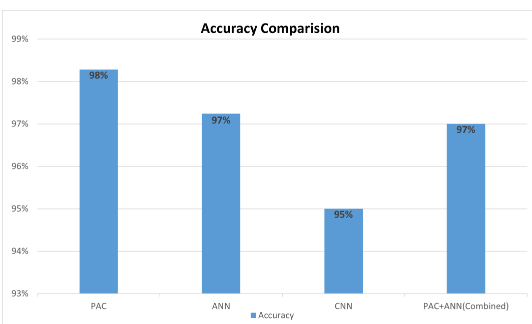

Fig. 4. Accuracy Comparison

In our study, we have employed a modified Passive Aggressive Classifier (PAC) that has been fine-tuned for optimal performance, resulting in a superior accuracy rate compared to the standard PAC. This meticulous optimization underscores our commitment to precision and the reliability of our fake news detection system. As depicted in Fig. 4. the accuracy comparison of different models is presented, with our enhanced PAC achieving a remarkable 98% accuracy, demonstrating its robust predictive capabilities and its proficiency in making accurate predictions on unfamiliar data. Additionally, the Artificial Neural Network (ANN) achieved an accuracy of 97.73%, and the combined PAC and ANN approach yielded an accuracy of 97.30%. The CNN model, included for comparative purposes, attained an accuracy of 97%, thereby affirming the effectiveness of our system across various methodologies.

In TABLE II, a comparison is made between the Passive Aggressive Classifier (PAC), Artificial Neural Network (ANN), CNN, and the combined model in terms of accuracy, precision, recall, and F1 Score:

TABLE II
Models Comparison Table

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**
PAC	98%	0.98	0.97	0.98
ANN	97%	0.98	0.96	0.97
CNN	95%	0.95	0.98	0.97
PAC+ANN	97.3%	0.96	0.98	0.97

Fig. 5. illustrates the confusion matrices for the Passive Aggressive Classifier (PAC), the Artificial Neural Network (ANN), and the combined PAC+ANN model. These matrices provide a visual representation of the performance of each model, detailing the number of correct and incorrect predictions made, thus offering insights into the true positive, false positive, true negative, and false negative rates of classification. Specifically, we observed a high number of true positives and true negatives, indicating the model's ability to correctly classify both positive and negative instances. However, we also noted some false positives and false negatives, albeit at a notably lower frequency. These errors suggest areas where the model may benefit from further refinement, potentially through feature engineering or fine-tuning of hyperparameters. Nonetheless, the overall performance, as indicated by the confusion matrix, underscores the efficacy of our model in accurately predicting class labels. Such insights gleaned from the confusion matrix serve as invaluable feedback for optimizing our model and enhancing its predictive capabilities in real-world applications.

V. Conclusion

The quest to identify and mitigate fake news is increasingly challenging due to the dynamic and sophisticated nature of misinformation, especially with the rise of AI-generated content. Our research contributes to this ongoing battle by introducing a hybrid detection model that integrates Passive Aggressive Classifiers and Artificial Neural Networks. This robust framework is trained on an extensive dataset, encompassing text, titles, and authorship, to accurately assess the authenticity of online articles. Through meticulous preprocessing, feature engineering, and dimensionality reduction, we significantly enhance the quality of our data.The incorporation of Rapid API's diverse services further empowers our model, enabling sentiment analysis, topic modeling, and rigorous fact-checking.

Looking ahead, our focus is on fortifying AI's discernment through advanced natural language processing and machine

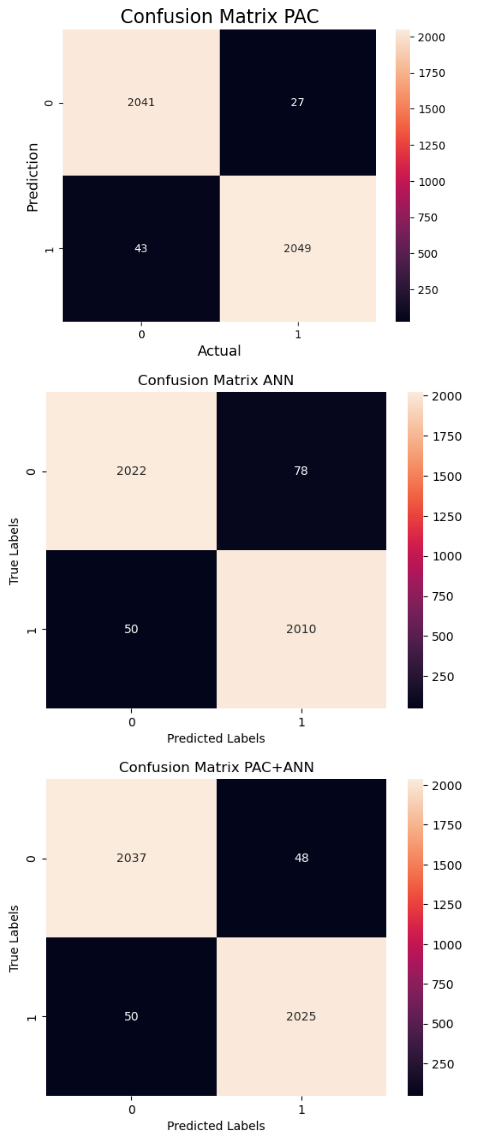

Fig. 5. Confusion Matrix Of PAC, ANN & PAC+ANN

learning techniques, particularly for multimedia content. The development of real-time verification tools is paramount, serving as a linchpin in the rapid authentication of news and curtailing the spread of misinformation. In this regard, we aim not only to enhance the accuracy of fake news detection but also to safeguard the integrity of journalism and elevate the quality of public discourse.

References

[1] Z. Qu, Y. Meng, G. Muhammad, and P. Tiwari, "Qmfnd: A quantum multimodal fusion-based fake news detection model for social media," *Information Fusion*, vol. 104, p. 102172, 2024.

[2] J. Golbeck, M. Mauriello, B. Auxier, K. H. Bhanushali, C. Bonk, M. A. Bouzaghrane, C. Buntain, R. Chanduka, P. Cheakalos, J. B. Everett *et al.*, "Fake news vs satire: A dataset and analysis," in *Proceedings of the 10th ACM conference on web science*, 2018, pp. 17–21.

[3] X. Fang, H. Wu, J. Jing, Y. Meng, B. Yu, H. Yu, and H. Zhang, "Nsep: Early fake news detection via news semantic environment perception," *Information Processing & Management*, vol. 61, no. 2, p. 103594, 2024.

[4] G. Pennycook and D. G. Rand, "The psychology of fake news," *Trends in cognitive sciences*, vol. 25, no. 5, pp. 388–402, 2021.

[5] P. Przybyla, "Capturing the style of fake news," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 01, 2020, pp. 490–497.

[6] D. M. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild *et al.*, "The science of fake news," *Science*, vol. 359, no. 6380, pp. 1094–1096, 2018.

[7] V. Klyuev, "Fake news filtering: Semantic approaches," in *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*. IEEE, 2018, pp. 9–15.

[8] R. Oshikawa, J. Qian, and W. Y. Wang, "A survey on natural language processing for fake news detection," *arXiv preprint arXiv:1811.00770*, 2018.

[9] M. C. de Souza, M. P. S. Gôlo, A. M. G. Jorge, E. C. F. de Amorim, R. N. T. Campos, R. M. Marcacini, and S. O. Rezende, "Keywords attention for fake news detection using few positive labels," *Information Sciences*, vol. 663, p. 120300, 2024.

[10] "kaggle fake news," https://www.kaggle.com/competitions/fake-news, 2014, accessed: Date accessed.

[11] H. Ahmed, I. Traore, and S. Saad, "Detecting opinion spams and fake news using text classification," *Security and Privacy*, vol. 1, no. 1, p. e9, 2018.

[12] K. Nakamura, S. Levy, and W. Y. Wang, "r/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection," *arXiv preprint arXiv:1911.03854*, 2019.

[13] W. Y. Wang, "" liar, liar pants on fire": A new benchmark dataset for fake news detection," *arXiv preprint arXiv:1705.00648*, 2017.

[14] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media," *Big data*, vol. 8, no. 3, pp. 171–188, 2020.

[15] A. Vlachos and S. Riedel, "Fact checking: Task definition and dataset construction," in *Proceedings of the ACL 2014 workshop on language technologies and computational social science*, 2014, pp. 18–22.

[16] H. Ahmed, I. Traore, and S. Saad, "Detection of online fake news using n-gram analysis and machine learning techniques," in *Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments: First International Conference, ISDDC 2017, Vancouver, BC, Canada, October 26-28, 2017, Proceedings 1*. Springer, 2017, pp. 127–138.

[17] E. Dai, Y. Sun, and S. Wang, "Ginger cannot cure cancer: Battling fake health news with a comprehensive data repository," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 14, 2020, pp. 853–862.

[18] I. Ahmad, M. Yousaf, S. Yousaf, and M. O. Ahmad, "Fake news detection using machine learning ensemble methods," *Complexity*, vol. 2020, pp. 1–11, 2020.

[19] J. Harsin, "Oxford research encyclopedia of communication," 2018.

[20] Á. Figueira and L. Oliveira, "The current state of fake news: challenges and opportunities," *Procedia computer science*, vol. 121, pp. 817–825, 2017.

[21] E. C. Tandoc Jr, "The facts of fake news: A research review," *Sociology Compass*, vol. 13, no. 9, p. e12724, 2019.

[22] J. P. Baptista and A. Gradim, "Understanding fake news consumption: A review," *Social Sciences*, vol. 9, no. 10, p. 185, 2020.

[23] M. Verstraete, J. R. Bambauer, and D. E. Bambauer, "Identifying and countering fake news," *Hastings LJ*, vol. 73, p. 821, 2022.