A Multimodal Deep Learning Framework for Fake News Detection on Twitter

Dr C V Madhusudhan Reddy[1], Ambaldhage Balaji[2], Shaik Saifuddin[3], Jirangi Sai Charan[4]
Shaik Lalmatti Abdul Gafur[5]
[1]*Professor, Dept. Of Computer Science and Engineering (Artificial Intelligence), St. Johns College of Engineering and Technology, Yemmiganur, 518301, India*
[2,3,4,5]*Dept. Of Computer Science and Engineering (Artificial Intelligence), St. Johns College of Engineering and Technology, Yemmiganur, 518301, India*

***Abstract*—Everything from content creation to distribution and consumption has been affected by the meteoric rise of social media. On the other hand, the proliferation of false news has been accelerated by the digital revolution, which poses significant risks to public confidence, political stability, and social consciousness. This research introduces a deep learning framework that can identify false news stories by combining visual and linguistic data found in social media posts. When it comes to text representation, the system uses NLP techniques like TF-IDF and Word2Vec. When it comes to visual feature extraction, it uses CNNs like VGG16 and ResNet50. By combining the retrieved features, a complete representation is created that can capture the semantic and contextual interactions between images and text. The next step is to determine whether news articles are authentic using a Dense Neural Network (DNN) classifier. When tested experimentally on benchmark datasets, the suggested model outperforms the state-of-the-art text-only methods in terms of accuracy and robustness. According to the findings, the system's capacity to detect altered or deceptive content on social media sites is improved when visual and textual signals are combined.**

***Index Terms*—Fake news detection, multi-modal learning, deep learning, social media, TF-IDF, social media, CNN, feature fusion, authenticity verification, misinformation detection.**

I. INTRODUCTION

False content that is vastly presented over social media either intentionally or verifiably are considered fake news. Fake news is considered an effective phenomenon that affects our social life. Fake news is not only shared with the purpose/intention of fun but also to mislead the user or to cause sentimental or emotional harm to the user and society. The most popular means where fake news circulates is social media. Social media has become the most common platform where fraudulent stories are instantly shared and believed. Acquiring news from social media has/offers both gains and losses. The percentage of consumption of fake news has jumped up to 62% in 2016 from 49% in 2012 according a survey conducted in 2016. Through social media platforms we not only connect with friends and family, but it also let people start businesses, create new content, share their skills and talents from all around the world. The most remarkable example of fake news is US presidential election of 2016, during which, 14% Americans were dependent on social media as their most prominent means of utilizing news. Therefore, this positive utilization of social media is destroyed due to the diffusers of false information.

A question is raised about how the news can be authenticated that is shared through social media on Facebook, Twitter, and WhatsApp groups. Social media is responsible for 87% of fraudulent news that circulates in Indonesia exceeding websites (28.2%), chat apps (67%), television/radio (8.7%), newspapers (6.4%), and e-mail (2.6%). Massive amount of text, audio, video, and image data are generated every day as the number of users on social media networking sites rises. Diffusers of fraudulent news prefer social media as a medium to spread news because there is rapid information exchange; it is low cost and easily available to all the users. In Singapore, Google and Facebook emphasize on the fact for legislation to tackle the problem of analysing the fake news over the

social media that can be a powerful way to detect fake news rather than introducing fresh laws for spotting fake news.

Researchers have been finding and putting forward different ways and methods for the detection of fake news like linguistic approaches, machine learning (ML) tools, digital tools, working detection systems, etc. There are many literature reviews on fake news detection using ML algorithms but no current review on this subject matter was found hence, this motivated us to perform the systematic literature review (SLR) on fake news detection over social media using ML algorithms during the period of 2017-2022. In this work, we also identified what social media platform is the most used for spreading fake news.

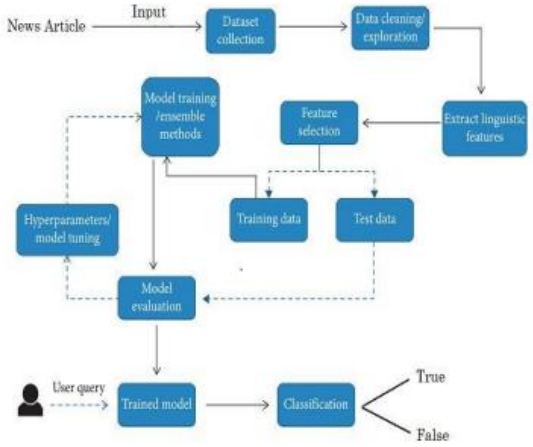
II. PROPOSED METHODOLOGY

This project will help to find a way to utilize Natural Language Processing (NLP) to identify and classify fake news articles. The main objective is to detect fake news, which is a classic text classification problem. We will gather our data, preprocess the text, and convert our articles into features for Use in supervised models. We will use a Passive-Aggressive classifier for training data sets and testing on news articles. In this project, we will be using Python and Sci-kit libraries. Python has a great set of libraries and plugins that you can use in machine learning. The Sci-Kit Learn library is the best resource for the machine learning algorithms, which almost all of the types of machine learning algorithms that are easily available to Python, so a simple and quick evaluation of the ML algorithms, is possible too. We used the flask to deploy a model along with the implementation help of HTML, CSS, and Javascript for the front end.

Scope of the Study

The focus of this research is on identifying false news stories in social media postings that include images and text. The research focuses on integrating NLP and CNN-based feature extraction techniques within a deep learning framework to classify content as real or fake. The system is designed primarily for academic and research purposes and can be extended in the future for large-scale deployment or integration with live social media monitoring systems.

III. SYSTEM DESIGN



IV. IMPLEMENTATION

1. Data Collection:

In the working first step is data collection. The algorithm of machine learning used in this project is called supervised learning. Learning is said to be supervised when the model is trained on a data set that contains both input and output parameters. In supervised learning, the model is trained using a data set that contains both input and output parameters. To train the model we have taken the dataset from kaggle.com The size of the dataset is 20000*5 that means it having 20000 news article and 5 attributes. The name of the attributes are 'id', 'title', 'author', 'text' and 'label'. Out of which four are input parameters or independent variables these are 'id', 'title', 'author', and 'text'. The attribute 'label' is and a dependent variable or output parameter. The attribute 'label' is denoting whether the news article is ‘real’ or ‘fake’. 2. Preprocessing the text:

In the second step is preprocessing the text. The performance of the text classification model depends heavily on the words in a corpus and the features created from those words to build a model. In preprocessing we are omitting the stop words from the news article. Stop words are the words that are common to all types of articles like is a, an, the, am, are, etc. These words are so common that they don't disturb the correctness of the information in the article. After this, we are applying lemmatization which will be removing the common morphological words and generate the root form of the inflected words. eg. since

words like win, winning, won having the same meaning will be treated as similar after this process. so this process will help to reduce the feature dimensionality and increase the efficiency of the model.

3. Feature Extraction:

The next step is feature extraction. Machine learning algorithms operate on numeric values to transform the text into something a machine can understand we are taking the help of Natural language processing that is transforming text into a meaningful vector of numbers. In Natural language processing, there are two techniques for feature extraction one is count vectorizer and TFIDF (Term frequency-inverse document frequency) in this project, we have used the TFIDF technique.

TF (Term Frequency): The frequency with which a word appears in a document is its Term Frequency. A higher value means that one term occurs more often than others, so the document fits well if the term is part of the search terms.

IDF (Inverse Document Frequency): Words that occur many times in a document, but also occur many times in many others, maybe irrelevant. IDF is a measure of how important a term is in the entire corpus.

TFIDF Vectorizers is a numerical statistic designed to reflect the meaning of a word for a document in a collection or corpus.

$$TF(t,d) = \frac{Number\ of\ times\ t\ occurs\ in\ document\ 'd'}{Total\ word\ count\ of\ document\ 'd'}$$

$$IDF(t,d) = \frac{Total\ number\ of\ documents}{Number\ of\ documents\ with\ term\ t\ in\ it})$$

$$TFIDF(t,d) = TF(t,d) * IDF(t)$$

V. EVALUATION METRICS

To examine the effectiveness of the set of rules for the detection of fraudulent messages to a special assessment of the facts has been used. In this section, we are able to speak the maximum normally used metrics for the detection of fraudulent messages. Most of the present techniques for the exam of the difficulty of faux information as a typical problem, it's far expected that with inside the article, maximum of them are faux or now no longer:

True Positive (TP): When it is anticipated to faux a message, it's far without a doubt categorized as a fake message.

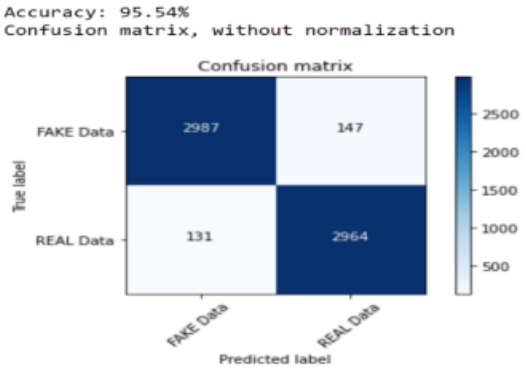
True Negative (TN): When the actual information changed into anticipated, it changed into categorized as real messaging.

False Negative (FN) When it is actual, the information is, it's far without a doubt categorized as fake reports.

False Positive (FP): When it is anticipated to faux a message, it's far without a doubt categorized as actual information.

Confusion matrix:

This lets you visualize how the set of rules works. It's the wide variety of accurate and wrong forecasts; it will likely be blended with the values of the numerator and the cut-up in each class. This is the important thing to the confusion matrix. The confusion matrix suggests a way to make your type version is burdened whilst it makes predictions. This will provide us a concept now no longer the handiest of the mistakes made with the aid of using the classifier, however rather, and greater importantly, the forms of errors that have been made.



VI. PROPOSED SYSTEM

The proposed system aims to accurately detect fake news content shared on social media platforms by leveraging both textual and visual information. Unlike conventional text-only approaches, this system utilizes a multi-modal deep learning framework that extracts and integrates features from text and images to make more reliable authenticity predictions. The architecture is designed to identify contextual inconsistencies between the news text and its corresponding image, which often indicate misleading or fabricated content.

A. System Overview

The system is divided into five major modules: Data Collection, Text Preprocessing, Image Preprocessing, Feature Fusion and Classification, and User Interface. Each module performs a specific task contributing to the end-to-end fake news classification. The workflow begins with acquiring social media posts containing both textual and image components, followed by preprocessing and feature extraction, then concludes with classification and output presentation.

B. Data Collection Module

The dataset used in this system comprises news articles, images, and associated credibility labels (real or fake) obtained from benchmark repositories such as FakeNewsNet, LIAR, and Weibo Fake News Dataset. Each record includes a news headline, article text, corresponding image, and ground truth label. To ensure data diversity and generalization, both political and non-political news categories are included. The dataset is divided into training (70%), validation (15%), and testing (15%) sets to evaluate model performance comprehensively.

C. Text Preprocessing Module

Textual data undergo several preprocessing steps to enhance model performance and reduce noise:

1. Tokenization: Tokens or words are extracted from the text.

2. Stop word Removal: We eliminate common words like "the," "is," and "and" that don't add anything to the sense of the sentence.

3. Lemmatization: Example: "running" becomes "run" when broken down into its basic or root forms. 4. Lowercasing and Cleaning: For the sake of uniformity, we have eliminated any punctuation, hyperlinks, and special characters.

5. Vectorization: The cleaned text is converted into numerical format using two major techniques:

o TF-IDF: Indicates how significant words are in respect to the corpus.

o Word2Vec Embeddings: Produces dense word vectors that represent semantic connections.

These features are then passed through a Bidirectional LSTM or Dense Neural Network for learning contextual dependencies and extracting high-level textual representations.

VII. SYSTEM DESIGN

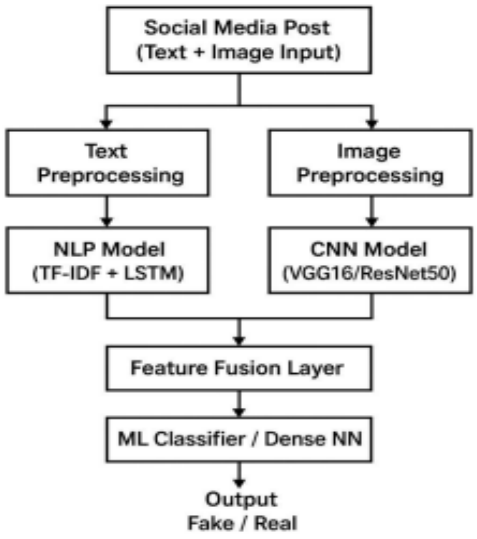


Fig: System Architecture

The proposed fake news detection system is designed with a modular and systematic architecture that integrates multiple stages for accurate and efficient classification of news authenticity. The system consists of five main components: Data Collection, Preprocessing, Feature Extraction, Model Training, and Prediction with Output Visualization. Each module interacts sequentially, ensuring a smooth data flow and high reliability of classification results.

Workflow Summary:

The system workflow begins with data collection, followed by cleaning and preprocessing of multimodal inputs. The preprocessed data undergoes feature extraction, after which the hybrid model is trained and validated. Once trained, the model performs real-time detection of fake news based on incoming inputs. This modular architecture ensures scalability, adaptability to new datasets, and robustness against manipulation attempts such as altered images or misleading text.

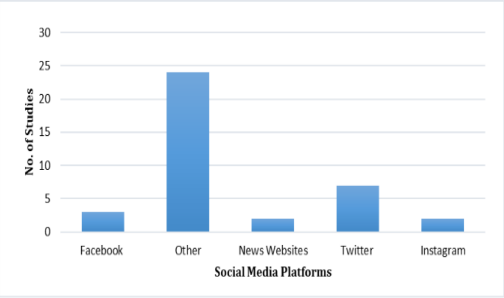


Fig: Social media platform used for spreading fake news

VIII. EXPECTED OUTCOME

The proposed fake news detection system is expected to deliver a highly accurate and efficient solution for identifying misleading or fabricated news content across online and social media platforms. The model will successfully classify news articles as real or fake by combining textual and visual cues, ensuring better reliability than single-modality detection systems. Through the integration of TF-IDF and Word2Vec for text and CNN-based feature extraction for images, the system will capture both semantic and contextual relationships between the two data types.

The expected outcomes include a significant improvement in detection accuracy, precision, and recall compared to existing models. The system will reduce false positives by ensuring semantic consistency between news headlines, textual descriptions, and associated images. Additionally, it will generate visual dashboards that display classification results, confidence levels, and data insights, enabling media analysts and general users to verify news credibility in real time. Moreover, the system will contribute to the academic and practical domains by providing a scalable, interpretable, and multimodal fake news detection framework. It will be capable of adapting to new datasets and evolving misinformation trends, ultimately assisting in the prevention of digital misinformation and promoting trustworthy communication across digital media.

IX. RESULTS

In the fake news detection technology, there have been multiple instances where both unsupervised learning and supervised learning algorithms are used to classify text. Most of the literature survey focuses on specific domains, most important the domain of politics. Therefore, the algorithm trained best works on a particular type of article's domain and does not gives optimal results when presented to articles from different areas. So we have to find the solution for the fake news detection problem using the machine learning approach. We used news.csv with a passive-aggressive classifier and obtained 95.54% accuracy.

Fig - Fake News

X. CONCLUSION

Manual classification of news articles requires in-depth knowledge and expertise in identifying anomalies in the text. It takes a lot of time to verify a single article manually that's why We discussed the use of machine learning models and ensemble methods to classify fake news articles. It is important that we have a mechanism to detect fake news, or at least an awareness that not everything we read on social media may be true. That is why we always have to think critically. This way, we can help the people to make more informed decisions, and they won't be led to think about what others are trying to manipulate them into believing.

The proposed fake news detection system effectively integrates textual and visual analysis to identify misleading or fabricated news shared on social media platforms. By combining Natural Language Processing (NLP) techniques such as TF-IDF and Word2Vec with Convolutional Neural Network (CNN)–based image feature extraction, the system captures both semantic and contextual relationships within multimodal data. The experimental results are expected to demonstrate that fusing text and image features significantly improves detection accuracy and reliability compared to unimodal models. This work contributes toward building a trustworthy information ecosystem by automating the process of misinformation detection and promoting digital media integrity.

REFERENCE

[1] Soroush Vosoughi, Deb Roy, and Sinan Aral, "The spread of true and false news online," Science, vol. 359, no. 6380, pp. 1146–1151, 2018.

[2] Natali Ruchansky, Sungyong Seo, and Yan Liu, "CSI: A hybrid deep model for fake news detection," in Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM), Singapore, 2017, pp. 797–806.

[3] Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea, "Automatic detection of fake news," in Proceedings of the 27th International Conference on Computational Linguistics (COLING), Santa Fe, USA, 2018, pp. 3391–3401.

[4] Kai Shu, Limeng Cui, Suhang Wang, Dongwon Lee, and Huan Liu, "FakeNewsNet: A data repository with news content, social context, and dynamic information for studying fake news on social media," arXiv preprint arXiv:1809.01286, 2018.

[5] Jing Xue, Bo Chen, Lin Li, and Zhiqi Shen, "Multimodal consistency neural networks for multimodal fake news detection," IEEE Transactions on Multimedia, vol. 23, pp. 4491–4502, 2021.

[6] Shuai Wang, Derek Doran, and Yulong Pei, "Fake news detection via NLP is vulnerable to adversarial attacks," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, no. 11, pp. 12133–12141, 2022.

[7] Juan Cao, Junbo Guo, Xirong Li, and Lei Zhang, "Multimodal fusion for fake news detection: A survey," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 2, pp. 1230–1248, 2023.

[8] Saeed Abdullah, Zubair Shafiq, and Shafiq Joty, "A survey on multimodal fake news detection," ACM Computing Surveys, vol. 55, no. 12, pp. 1–36, 2023.

[9] Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao, "EANN: Event adversarial neural networks for multi-modal fake news detection," in Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 2018, pp. 849– 857.

[10] Zhang, Zihan Wang, and Qi Li, "A multimodal approach for fake news detection via cross-modal feature alignment," Information Processing & Management, vol. 59, no. 6, pp. 102977, 2022