Multimodal Fake News Detection: Integrating OCR and Deep Learning Models for Text and Image Analysis

MSc Research Project

MSc Data Analytics

Jayaram Vennapu
Student ID: X22219978

School of Computing

National College of Ireland

Supervisor: Shubham Subhnil

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	……. ……………Jayaram Vennapu…………………………………………………………		
Student ID:	…………………x22219978……………………………………………………………..……		
Programme:	…………Data Analytics……………………	**Year:**	…2024-2025..
Module:	………………Research Project…………………………………………….………		
Supervisor:	………………Shubham Subhnil…………………………………………………….………		
Submission Due Date:	………………12/12/2024……………………………………………………….………		
Project Title:	Multimodal Fake News Detection: Integrating OCR and Deep Learning Models for Text and Image Analysis		
Word Count:	…………8320…………	**Page Count**	…………19……………

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: …………………Jayaram Vennapu…………………………………………………

Date: ………………………………12/12/2024……………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□NA
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□NA
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□NA

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Multimodal Fake News Detection: Integrating OCR and Deep Learning Models for Text and Image Analysis

Jayaram Vennapu

X22219978

Abstract

Fake news identification is an essential task of the current age, especially with the help of multimedia channels. This work proposed a dual Optical Character Recognition (OCR) method supplemented by a multimodal deep learning model to distinguish between real and fake news. The dataset comprises images with raw textual news content, and it is divided into a training set with 88%, a validation set with 8%, and a test set with only 4%. Data augmentation is done by rotating the images within the plane, changing saturation and exposure, making random cuts to the image, and resizing the images to be 640 * 640 pixels. For text analysis, the OCR used is OCR 2.5 and the information extracted from images are tokens using a BERT-base-uncased tokenizer. The attached text is pre-trained and fine-tuned on a Bert Bert-based transformer model for three epochs, test accuracy of 95.12% and an F1-score is 95.12%. For image analysis, a fully connected neural network, CNN-based, ResNet-18 is fine-tuned from the pre-trained ImageNet model and used to classify images achieving a test accuracy of 97.56 percent, and a test F1 score of 97.56 percent. The measures of accuracy, precision, recall, and F1-score prove the efficiency of the proposed system. This general approach unites text and image recognition, using OCR and deep learning innovations for recognizing fake news, providing a stable solution with top results.

Keywords: Fake News Detection, Optical Character Recognition (OCR), BERT-based Transformer, CNN-based ResNet, Image Classification.

1 Introduction

1.1 Background

Fake News Detection involves the recognition and removal of fake, untrue, or forged information posted on news sites and social media Sahoo and Gupta (2021). Fake news hurts public opinion and decision-making, as well as electoral processes, which makes this phenomenon's detection an important area of research. While the internet brings in more and more sources to encourage new-age comparatively, most content that spreads becomes viral very quickly without consideration of the fact's accuracy. Thus, fake news uses appeals to emotions, selection of the news source itself as well as the language, which makes it hard to verify manually Paschen (2020). Fake news detection research background is based on the increase in the requirement to fight against misinformation, especially after specific events that were held and influenced by false information, including political elections and pandemics. Earlier techniques involved the verification of facts by manual searching Bento et al. (2020), while with the increase in content production, these techniques proved to be inadequate. Due to integrated NLP and computer vision in the recent past, today's fake news detection systems can process both text as well as media.

Current methods include fine-tuned BERT transformers, which cover textual content analyzing linguistic dependencies, contextual information, and sentiment; ResNet derived from CNNs, identifies changed visuals or contextually anomalous inserts. Other tools like Optical Character Recognition (OCR) when integrated improve detection through extracting and analyzing of texts included within images. However, these factors such as language barriers, cultural differences, and other sophisticated fake content generation methods such as deepfakes remain to be hard to crack. These challenges therefore can only be met through teamwork involving technology experts, educators, and policy makers to work on the best ways of reducing the effects of fake news when disseminating information.

1.2 Aim of the study

The aim of this work is to improve the methods of identifying fake news through the application of innovative technologies in text and image processing. The specific aim is to build a strong model that will accurately point out, whether the presented text and images are fake or real news. The research utilizes Optical Character Recognition (OCR) to recognize text from images which can include screenshots or graphical representations of an article that is fake news. The text extracted is then preprocessed and utilized to train a fine-tuned transformer-based BERT model for efficient news categorization. Moreover, a Convolutional Neural Network (CNN) approach is incorporated by using ResNet model to identify other hidden visuals of an image that may represent a fake news. The idea of the proposed study is to combine the application of OCR for textual analysis and a CNN for image analysis in an attempt to offer a twofold solution to the case of fake news identification. The quantitative performance of the proposed model is analyzed based on various performance parameters including accuracy, precision, recall, and F1-score. In the end, this study aims at assisting in proposing better ways of confronting fake news incidences liberating modern media.

1.3 Research Objectives

There are two research objectives of this study which are as follows:

1. To integrate OCR technology for text extraction from images: This objective focuses on applying OCR 2.5 to process visual content such as images and screenshots commonly associated with fake news articles, extracting embedded text for further analysis.
2. To fine-tune a BERT-based uncased transformer model for fake news detection: The study aims to fine-tune a pre-trained BERT model using the extracted text from OCR to classify news content as either real or fake, utilizing the power of attention mechanisms to enhance accuracy in text-based analysis.

1.4 Research Questions

There are some research questions of this study which are as follows:

1. How effective is OCR 2.5 in extracting relevant textual content from images in news articles, and how does this impact the performance of fake news detection models?
2. What is the comparative performance of a fine-tuned BERT-based uncased transformer model in detecting fake news when applied to both text extracted via OCR and traditional text-based datasets?

2 Related Work

2.1 Traditional Approaches to Fake News Detection

Previous research on fake news detection employed rule-based methodologies of text classification together with pattern recognition. The rule-based systems use IF-THEN rules to categorize content, and often use dictionaries or structured patterns in performing the linguistic analysis, whiles machine learning models like convolutional neural networks (CNNs) to solve tasks like hashtag recommendation and semantic analysis. A major objective of these approaches is to detect falsehoods in large data feeds, for example from multiple social media feeds or curated news articles. That’s why while these approaches are efficient in organized environments they can be hardly scaled and applied to the changing language patterns and many-to-many communication scenarios.

The work described in Alotaibi and Alhammad (2022) seeks to identify and classify Arabic fake news tweets on COVID-19 to mitigate the negative consequences of fake news. A rule-based system by the authors used an Arabic dictionary for text classification to analyze a data set of 5 million tweets. The model categorizes fake news into six categories: there are six broad genres; entertainment, health, politics, religious, social and sport. One of the difficulties was developing the rule-based model for the Arabic language, which includes so many references and complex patterns to detect. The model obtained 78.1% of accuracy, 70% of precision, and 98% of recall identifying more than 26,000 of fake news tweets. It pointed out that there was a decline in the generation of fake news over time as the public became more educated, adding that the social category was relevant in most Arab countries, while the entertainment category had the lowest dissemination. But the rule-based approach may be less scalable and flexible to capture dynamic distribution of the language and strictly following rules may not identify contextual or new emerging fake news.

The work done by Kalaiselvi (2020) provides a solution to the question of enhancing the hashtag recommendation for breaking news as well as streams by explaining a rule-based methodology for live incorporation of news and social streams. As for the contributions of this work, there are richer techniques for data collection and feature computing introduced for enlistment feature in addition to the existing systems built upon the CNNs for hashtag recommendation of NL processing tasks. Discussed next is the use of IF-THEN rules in classifying data and identifying the relationship between Twitter hashtags and the news media, critical proximity for online users. The approach was intended to increase the effectiveness of hashtag suggestions, prompting they specifically targeted the breaking news moments and to optimize the integration of news content with social platforms. However, the use of rule-based classification also has some disadvantages against flexibility when facing different types of language and dynamical trends, and the model's performance on various circumstance could be influenced. Nevertheless, the study provides a viable solution to a practical problem of organizing real-time information flows adequately.

2.2 Machine Learning Approaches to Fake News Detection

There are some studies which is related to Fake News detection using machine learning models are Decision Trees (DT), Support Vector Machines (SVM), Naïve Bayes (semi-naïve Bayes + Gaussian Naïve Bayes), Passive Aggressive Classifiers. These approaches rely on feature engineering methods like Term Frequency-Inverse Document Frequency (TF-IDF), not only but engulf n-gram (unigram, bigram, trigram) too which usually conjugates with feature selection like Information Gain. Although overall there have been improvements noted in DT models, the SVM and other classifiers also give very acceptable performances subject to the dataset and the type of test applied. However, issues such as a limit on the source of the dataset and its generalization are still a primary issue that indicates further comprehensive studies for better performance and functional validation.

The first work is done by Shaikh and Patil (2020) who has paid much of the attention to solve the issue related to fake news which has significant impact on different field such as political and educational system. The work presents the developed solution in the form of a fake news detection model with SVM, Naïve Bayes, and Passive Aggressive Classifier utilizing a feature extraction process through Term Frequency-Inverted Document Frequency (TF-IDF). Despite the availability of datasets not being very high, the model gave a high accuracy of 95.05 % through SVM classifier by considering tf-idf features which proved the efficiency of the model in classifying fake news. However, this study is limited by the use of constrained datasets; this means that, when tested on other data, differently structured data, the accuracy of the model has dropped sharply. This study highlights the significance of distinguishing fake news for real and reiterates the viability of classification-centered techniques as a method of doing so, even if the accessibility of more expansive data and sustained improvements may improve the model.

The second study is given by Baarir and Djeffal (2021) study where the author discusses the concern of fake news production due to the advancement of the communication technologies and social media. This paper suggests an approach to fake news detection using machine learning with TF-IDF of bag-of-words and n-grams as extraction features and SVM as a classifier. In an effort to supplement the availability of data, the authors also suggest a new dataset comprising of fake and true news for training the model. The achieved metrics show that the system has high accuracy in identifying fake news, revealing the applicability of the approach. However, since the present study focuses only on a custom dataset, it may not be easily extended to various other news domains. The enlargement of the dataset diversity or the attempt to consider other classifiers might strengthen the proposed system and widen the sphere of its utility.

The study which is provided by Krishna et al. (2022) attempts to increase the realness of fake news detection on social media by comparing the results of the Decision Tree (DT) and Support Vector Machine (SVM) machine learning algorithms. Following a direct comparison of the two algorithms with an equal number of samples (N=311 each), accuracy and precision measurements were determined by the researchers. The outcomes found that the DT algorithm was higher than the SVM algorithm, with an accuracy of 97.67% and 91.74%, respectively and at 95% CI there were significant differences found in an accuracy (Md

=3.893; p=0.092) and precision (Md= 0.593; p=0.825). The analysis also shows that the DT algorithm has better performance over fake news detection in social media. Nevertheless, the complexity of the approach is confined by the project's dataset size and uniqueness, which might have a certain impact on the generality of the obtained results in a broader and non-homogeneous dataset or in the real environment. Future studies can work with even larger and different samples to confirm the findings.

Finally, Irena and Setiawan (2020), give a solution to the identification of hoax news on Twitter by using the Decision Tree C4.5 classification on a dataset of 50610 tweets. Unlike previous studies, it incorporates multiple test scenarios: The different ones included, simple classification, classification with weighting features using the TF-IDF frequencies, and features that were selected using Information Gain together with the weights. Generation of features includes the use of unigram, bigram and trigram n-grams. The study found that combining weighting features and feature selection yielded the best accuracy of 72.91% using a 90:10:0 ratio of training-to-test data with the computational model flexible of 5,000 unigram features. Although the approach presented shows how the feature weighting and selection can be used to improve the classification, the overall accuracy achieved call into question the generality of the cover model. This may have been a result of the inherent nature of the Twitter data, class imbalance or the fact that the fixed-size feature vector retains the first 10,000 features only. It therefore leaves future work to investigate broader feature space, different algorithms, and a wider set of instances for better detection performance.

2.3 Deep Learning Approaches to Fake News Detection

For fake news detection, some deep Learning techniques have been considered and various models have been introduced for better fake news detection. Current primary used models are Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRUs), and Concatenated CNNs (C-CNNs). CNNs and C-CNNs are useful in feature extraction and classification application and LSTMs and GRUs work well with the sequence data. Furthermore, existing word representation methods such as GLOVe are employed for embedding the text data. These models have brought out better performance thereby providing an accuracy as high as 99.6% The deep learning approaches are seen to improve the performance of the algorithm in the detection of fake news on social media platforms.

The study Kaliyar et al. (2020) developed a deep convolutional neural network known as FNDNet for automatically detecting fake news, especially on social media and web-based forums, where fake news has resulted in the erosion of the public's confidence in the media. Different from handcrafting of the feature, FNDNet aims to locate the discriminative features for fake news classification through multiple layers of the deep neural network. The method uses a deep Convolutional Neural Network (CNN) structure and different features are obtained from different layers of the CNN and the performance of the method is compared with other baselines. The proposed model was trained and tested on benchmarked datasets and the accuracy of our test data reached 98, 36% which shows the model has good generalization. The performance evaluation measures such as Wilcoxon, false positive rate, true negative rate, precision, recall, F1 score as well as accuracy proved that the proposed model yielded promising results in detecting fake news relative to the existing approaches. However, some concern arises because limited explorations into the scalability of the proposed model is made to possess better compatibility in handling larger and varied data sets or tests the model in different languages and cultural settings that may affect its usability in a real-world application.

The study Abualigah et al. (2024) suggests an enhanced method for enhancing the fake news detection systems given the increasing problem of spam and malicious accounts generating and or promoting fake news over social media platforms including popular ones such as Facebook and twitters. This work focuses on improving the preprocessing stage of fake news detection, especially by adopting GloVe (Global Vectors for Word Representation), which is an unsupervised learning model from Stanford University. GloVe produces word embeddings by considering globally distributed word co-occurrence matrices to learn word analogies from statistical data. The proposed method incorporates deep learning algorithms, CNN, and DNN along with LSTM networks and RNN for sequential data analysis. The model being tested used the Curpos fake news dataset and recorded a 98.974% accuracy. However, one of the possible weaknesses of this work is that the model is not analyzed with regards to its generalization capability in extended datasets and to its real-time efficiency as the density and heterogeneity of available data can present new difficulties in the generalization and computational domains.

Last, of all, the study by Sedik et al. (2022) introduces using a Deep Learning (DL) technique for the classification of Fake News on social media platforms which are now a major source of news but might be also used as a tool for sharing fake information that may harm the reputation and user trust. The proposed system consists of three phases: text coding, data characteristics and categorization. The text encoding phase uses Global Vectors for Word Representation (GLOVe) which encodes words and then embeds them to a fixed form for further processing. Four different DL models are proposed in the study; these models include CNNs, C-CNNs, LSTM, and GRU, and the objective is to ascertain which model best detects FNs. All the models were trained on the FNs and FNC datasets available on Kaggle and C-CNNs hit an impressive 99.6% accuracy with faster training than the other models. The effectiveness of this approach was assessed concerning a host of measures including precision, recall, F1 score, and accuracy, all of which outperformed previous models. However, some issues that cannot be answered in the study are the ability of the models to deal with bigger and more heterogeneous data sets or indeed the ability to implement the models in dynamic social media space where the characteristics of fake news can easily transform over a short period.

Table 1: Comparison of Different Techniques for Fake News Detection and Their Capabilities

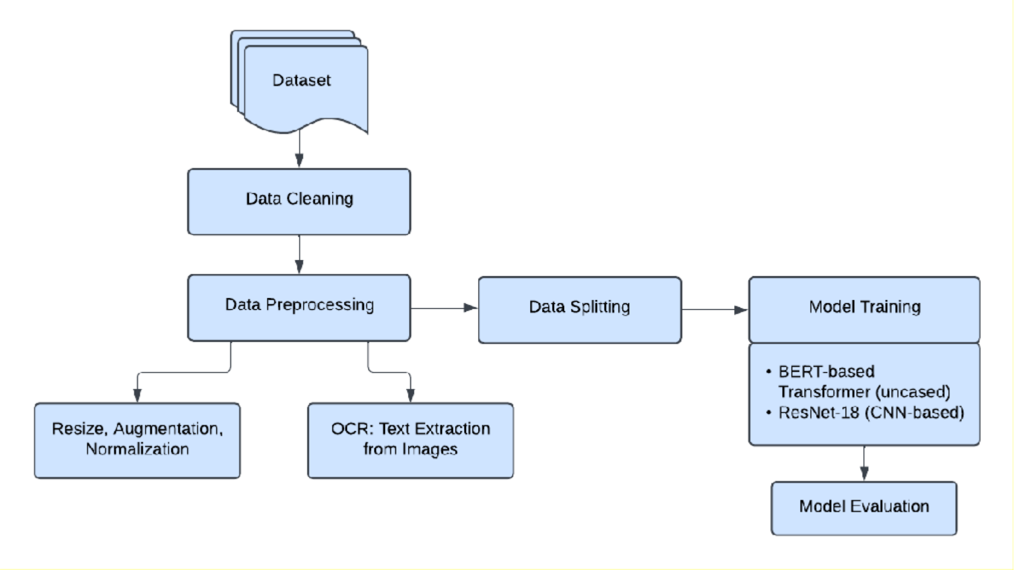
Technique	Can Handle Dynamic Language Patterns	Can Classify Fake News Across Multiple Domains	Scalable for Large Datasets	Can Handle Multilingual Data	Can Identify Emerging Fake News	Real-Time Detection Capability
Rule-Based Systems (Alotaibi & Alhammad, 2022)	+		+	+	+	+
Hashtag Recommendation (Kalaiselvi, 2020)	+			+	+	
SVM, Naïve Bayes, Passive Aggressive Classifier (Shaikh & Patil, 2020)			+	+	+	+
Decision Tree & SVM (Krishna et al., 2022)			+	+	+	+
C4.5 Decision Tree (Irena & Setiawan, 2020)	+			+	+	+
FNDNet (Kaliyar et al., 2020)				+		+
Deep Learning (CNN, LSTM, GRU) (Abualigah et al., 2024)						
Deep Learning with GloVe (Sedik et al., 2022)						

3 Methodology

3.1 Libraries Imported

This study combines several types of libraries to provide effective solutions for text and image processing in a broad array of libraries. Built with PyTorch, deep learning has its core components for defining neural networks with torch.nn, optimization methods with torch.optim, and data loading with torch.utils.data for controlling the batch management of the datasets. Torchvision is used for data augmentation and preprocessing of the images that may be used for OCR, with features such as resizing, rotation, normalization. OCR utilizes the Hugging Face Transformers library, including AutoTokenizer and AutoModel , to preprocess and fine-tune a BERT based uncased transformer model for the project. This library provides the Trainer and TrainingArguments APIs that make fine-tuning not so complex because it is easy to modify the hyperparameters that are used in the fine-tuning process.Trainer Arguments API is an important part of the scikit-learn’s metric computation tools: accuracy, precision, recall, and F1 answer functions are used for the exact measurement of the model’s performance. The Pillow (PIL) module is used for image processing, which allows operations such as resizing; an analysis of the augmentation effects is demonstrated. Support of techniques and techniques of visualizing results and processes is provided by the matplotlib tool – a universal tool for plotting. Because the default warnings often interfere with necessary ones and clutter the output, the warnings module and Python’s logging library are employed, with specific settings for controlling logs coming from transformers and other procedures. Also, some key sections of python such as OS library are integrated to facilitate directory management hence loading and saving datasets and models. When these libraries are combined, the project guarantees the successful passing of text and images, strong training, and stable validation of a model to develop an efficient system to detect fake news with better accuracy and in larger volumes.

Block Diagram



3.2 Data Cleaning

Data cleaning is an important part of the process where input data is checked and made clean. It includes the processes of stripping out other unwanted characters including symbol characters, spaces, and any other alphanumeric character that may occur due to the process of OCR text extraction. Text is also normalized to lower case in order to be consistent with cased BERT model. Also, stop words such as urls are omitted and any messy text such as wrong spelling is eliminated. This cleaning process ensures that there is cleaner data, which when extracted, is good for data tokenization and enhances BERT's effectiveness for fake news detection.

3.3 Data Preprocessing

In the context of this study, the first step related to data preprocessing in the framework of the "Optimizing Fake News Detection" project is focused on preparing the images containing the data for OCR analysis and subsequent text-based processing. First, all photographs are automatically rotated to planar position based on Fourier features and then scaled to 640×640 pixels with stretching applied uniformly to all images due to differences in the facial length of all members within a family. For higher model's stability and lower overfitting, data augmentation methods are used, creating three copies in regard to each image. These augmentations include random rotation within the range of -15° to +15°; saturation with the range of -25% and +25% and exposure with the range -25% and +25%; cutout augmentations which include three occlusion boxes that cover 29% of image area. Subsequently, OCR language detection is performed using OCR 2.5 an advanced model from the Hugging Face. The extracted text is then tokenized using the AutoTokenizer from the transformers version by Hugging face, transforming it into subwords while generating attention masks to handle scenario of padding properly. To make it compatible with the BERT based uncased transformer model labels and encodings are developed here for this purpose. Labels the input tokens, the attention structures and other aspects of the model input data make it possible to fine-tune the model in an integrated environment that would enable the model to progress from image-to-text stop-over on to fake-news classification.

3.4 Text Tokenization and Encoding

This study has been used the text tokenization and encoding procedures form part of the preparatory method of the extracted text that feeds into the fine-tuning of the BERT-based uncased transformer model. Following the tokenizer of the AutoTokenizer from Hugging face, the extracted text is preprocessed into subword tokens in order to address issues of out-of-vocabulary words. Since these subwords are the building blocks of words, the required flexibility in tokenization used in this case guarantees that every semantically distinguishable word is represented in a form that can be correctly mapped to the model's vocabulary, including rare words and compounds. In addition to tokenization, attention masks are created to distinguish padding within the sequence, so that the model is only paying attention to all the meaningful parts of the input text and ignoring the rest of it. They are very crucial in such training because they help in achieving optimal computational efficiency. Moreover, the tokenizer maps all the tokens into numerical representations that the model can understand as well as attaches these numerical representations to the labels which are required for the supervised learning. The labels are designed with regard to the classification goal of fake news identification. Although it doesn't necessarily have to be this way, this integrated approach to tokenizing and encoding not only normalizes the input data, but also prepares it in the best possible way for the fine-tuning process and for BERT model at the conclusion of this project.

3.5 Dataset Description

The dataset that was employed in the "Optimizing Fake News Detection" project includes images with text where such text as an article or headline is intended for OCR extraction. The dataset is split into three parts: Training set 1812 images, 88%, validation set 174 images, 8% testing set 82 images, 4%. Some of the preprocessing of images include orienting it automatically and resizing the images to a certain dimension of 640*640 pixels while using the stretch option. To improve model reliability, data augmentation approaches are used to produce three different forms of every training image. These augmentations include random rotations restricted to -15 to + [1]15 degrees, saturation values between -25% and +25%, brightness variability of the same range, and cutout augmentation of three random masks per image, each taking 29% of the image size. First, these images will be leveraged to OCR 2.5, a model developed based on general OCR theory, and then the processed text will be again run through AutoTokenizer and AutoModel from Hugging Face transformer library. The BERT-based uncased transformer model is then fine-tuned by this processed text. Preprocessing involves splitting the dataset into three key components: labels and encodings on this view, each is optimized and structured through a tokenizer to fit the BERT model attention mechanism. Here, the fine tuning is performed over three training epochs known as tiers and the training argumentation to improve the performance is unique. The framework entails the computation of metric after training to determine the reliability of the model developed. This makes it possible to detect fakes news through synergy between OCR and next generation natural language processing strategies on the extracted text.

3.6 Data Visualization

In the work Figure 1 presents three samples of real images as to various sources such as the football game, the horses in the American west and the motion picture scene. The goal of this figure is to present examples of realistic image material which might be employed in the scenario of fake news identification study. Thus, it is possible to understand how characteristics and content of these real images can be used to define approaches to distinguish them from fake or manipulated images that might be used as misinformation. One of the steps now necessary and transformative in designing algorithms and models that can detect fake news, which is now a rampant problem, is comprehending the physical signs and contexts contained in authentic images. This figure can be used as reference for researchers towards real images and distinction from fake or manipulated visual content that are useful toward improving fake news detection systems.

[1] https://universe.roboflow.com/carlo-almeda/fake-news-image-classifier/dataset/1

Figure 1: Sample Real Images (Original)

A set of the fake images is shown in the Figure 2, where one can easily distinguish between real and fake images, which can be used to track the further spread of fake images and study the manipulative visual content. Most of these images seem fake or potentially edited, especially in a way to disseminate fake news or fake information. Examining the features, signs, and context found in these fake images proves important for researchers and developers who are charged with the task of designing fake news identification systems. trying to comprehend the manner in which such content is produced and distributed, they are able to develop improved procedure and formulas for identifying photorealistic false or misleading visual information. This figure is a direct counterpart to what was described in Figure 1, and it consists of an array of examples that can be used for testing and training of fake news identification tools to make information sharing platforms more reliable and trustworthy in th[1]e future.

Figure 2: Sample Fake Images (Original)

Figure 3 shows the sample of real images from Figure 1 with the resize transformation applied to them. These pictures have been normalized for size to a resolution of 224x224 pixels that is commonly used by most models of machine learning in computer vision. This action is important because it means that all images that are to be analyzed by those algorithms which are aimed at fake news detection will be of the right size to be properly processed and compared. This transformation is vital and the common property ensures that the models learn from the images and do not consider image size or aspect ratio. Using this resize transformation, the resultant 224*224 images keep a relatively clear view of the original content as well as meet the requirements of the targeted machine learning models. This step is also a preprocessing step very crucial in the creation and subsequent fine tuning of fake news detection systems, since failure to properly preprocess the input data is likely to lead to skewed results and consequent low reliability of the system.

Figure 3: Resized Images (224x224)

Figure 4 also shows the application of a horizontal flip transformation to the sample real images used in figure 1. Vertical flipping has been applied to the images, so they were rotated along the vertical axis by 180 degrees. It is frequently utilized in machine learning preprocessing techniques to improve the data variety by adding extra samples while changing the fundamental information of the image. When the model sees the same image in normal and mirrored form it will be able to recognize the appropriate features of the image more effectively enhancing the general performance and validation of the model. Taken in the overall framework of fake news detection, the horizontal flip transformation can be a good way for learning possibilities of manipulations with images

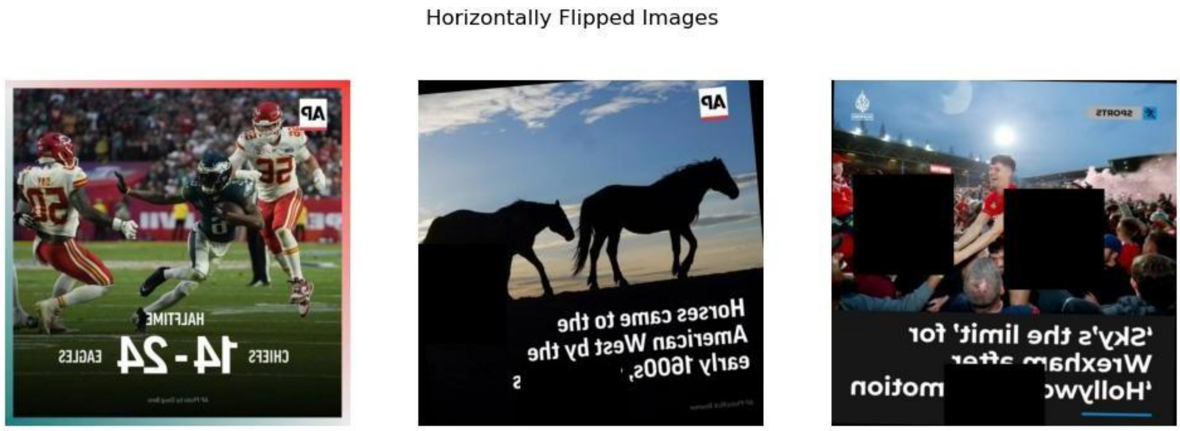


Figure 4: Horizontally Flipped Images

As illustrated in figure 5, the original images in pixel format are transformed into tensors for learning tasks. It is an important process of the data transformation towards the usage of the ML models since most deep learning frameworks as well as the majority of the algorithms shall expect input data to be in the tensor form. Tensor is a general concept for a collection of numbers that provides an efficient way to store all the values that might be assigned to the pixel of an image. The images are also transformed into tensor format to enable easy feeding of the data into computational pipelines of neural networks, and other machine learning models. This process, sometimes, requires scaling of the pixel values, often to a range of 0 to 1 since it's most convenient for some models to work at this range of inputs data.

Figure 5: Converted to Tensor

Figure 6 shows the normalisation transformation of the sample real images used in this study. Normalization as a part of data preprocessing steps is used to scale the pixel values within the image to a standard range of scale 0 to 1. This is a crucial transformation to make so as to ensure that the input data lies within the same scale and the scale makes other operational scales possible for other machine learning algorithms. Such normalization of the pixel values allows the models to pay attention to the variations in pixel values relative to each other rather than the overall variation of the pixel values. This makes the model perform better than the standard sigmoid function since it will improve the possibilities to generalize later on. Normalization can be a useful step in the data preparation process in the case of the fake news detection. Since the models need to compare pixel values and analyze pattern and features, it is beneficial to make all images have the same pixel value so that the pixel value does not influence the models. This makes the models less sensitive to the changes in the raw pixel data, thus making it easier for the models to detect features that separate the real form manipulated or fakes visual contents.

Figure 6: Normalized Images

3.7 Data Splitting

For this study, the dataset is reasonably distributed purposefully into three parts: training, validation and test sets. The training set contains 88% of data sets and provides 1812 images to train the model properly and to make the model identify similarities. This subset is to make ensure that the model has enough data to train for association between the features of the text and the fake news labels. Validation set is used, which constituted 8% of the data and consists of 174 images; the distribution of images is used to tune the hyperparameters of the model during the training process. This helps make the model generalised in that it should not be overly complex and fine-tuned to the model training data set. Last but not the least, the test set is with 4% which contains 82 images, and it is exclusively used to rate the model on input unseen data. This split helps to have a strong input, describing the effectiveness of the model when implemented practically. Both subsets have been selected in such a way that there are equal proportions of fake and real news in the same set to train and check the model on a feasible cross section type of the problem domain. This systematic splitting strategy allows for the assessment of the model and its improvement in the different development stages.

4 Design Specification

This workflow diagram in Figure 8 depicts a system of recognizing the fake news where both textual and visual information are integrated. It starts with Input Images from which data is obtained through Preprocessing. During the preprocessing, the images will go through auto-orientation, resizing and augmentation of the images to one that is suitable for the other tasks. The preprocessed data is then fed into two parallel processing streams: The two common utilizations include the Optical Character Recognition Extraction and the Image Feature Extraction. Using optical character recognition techniques, the OCR Extraction employs an identification and recognition of the text from images. This extracted text is then forwarded to the BERT Text Model which is a deep learning language model specifically fine-tuned for the purpose of classifying fake news through text. In parallel, a module for 'Image Feature Extraction' involves a ResNet CNN Model to decode the visible signals and structures in the images. That is why this CNN model learns to recognize visual properties when a particular news outlet may publish fake news. The textual as well as the visual features are then integrated in the Model Fusion, which produces a more general assessment by using BERT Text Model and ResNet CNN Model. Then, a Fake News Classification using the fused model is derived to classify whether the input content is indeed a Fake News or Real content. Then the system results are presented with relevant measures as key indicators for evaluating the validity of fake news identification. Since the proposed fake news identification task is a complex one involving natural language processing, computer vision, and multimodal fusion this workflow is ideal to solve the problem, and it will be almost impossible to break.

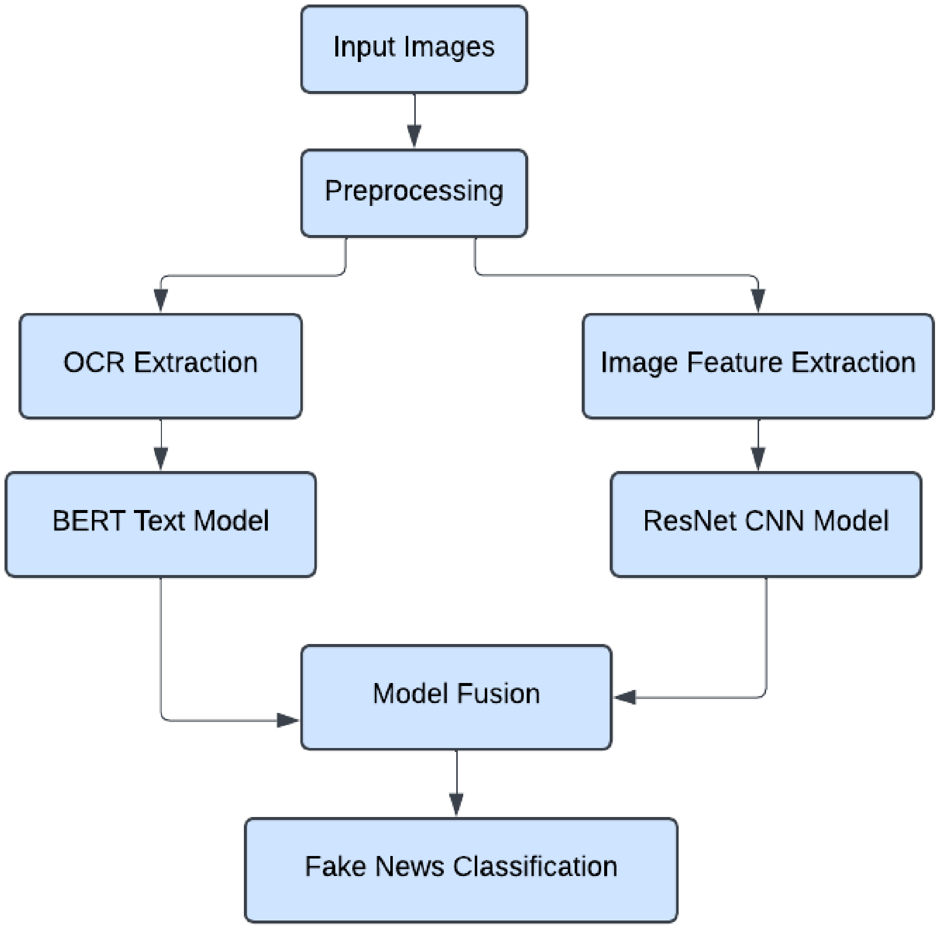


Figure 7: Proposed Workflow Diagram

5 Implementation

5.1 Tools and Technologies

In the present work, the methods of this study aimed at efficient and accurate textual and visual data analysis, involve a wide range of effective tools and technologies. For Optical Character Recognition (OCR), the project adopts the pre-trained OCR 2.5 model from Hugging Face as one of the best OCR models in the market for fine-grained image text extraction. The extracted text is in the unprocessed form and has to go through tokenization, encoding and compatibility with the fine-tuned BERT-based uncased transformer model through the help of the AutoTokenizer and AutoModel modules of the Hugging Face Transformers library. The fine-tuning process is complexly controlled by Trainer and TrainingArguments APIs to provide the easy and efficient training flows with tunable hyperparameters. Validation measurements like, accuracy, precision, recall, and F1-score are evaluated by using the scikit-learn function for the better understanding of the model's performance. On the image processing segment, arrangements such as resizing, rotation, and augmentation of the datasets are implemented for the PyTorch and Torchvision libraries. The data loaders are implemented through PyTorch for batch processing output. Data preprocessing steps involve using the Pillow (PIL) library to handle images and perform augmentation on images as part of visualization. Specifically, the torch.nn and torch.optim packages are used in constructing and adjusting bespoke neuronal network models; the matplotlib package is used to visualize data helping to recognize patterns and outcomes. Further, for fast model learning and testing, PyTorch utilizing GPU is incorporated as well into this work. Both logging and warning are handled using python modules where logging is done using Python standard library and the warnings. The project environment is well set with sophisticated Python libraries for efficient integration of OCR, NLP and machine learning to show the coordination of well-developed tools and sound approaches in developing an efficient fake news detection system.

5.2 Implementation of Text Analysis Model (Fine-Tuned BERT-Based Transformer)

The text analysis part of the fake news detection system includes a binary classification fine-tuned BERT-based transformer model. The implementation process first imports the BERT-base-uncase version of the tokenizer and model from the Hugging Face organization to include a tailored classification head for two output classes. Tokenizing of the text data in the dataset is done using the AutoTokenizer with padding and truncation implemented for proper sequences length. Tokenization provides tensors which would meet the model's specifications. The tokenized data is then mapped to PyTorch datasets using a custom FakeNews Dataset class that facilitates access to encodings and labels during training and assessment. Training, validation, and test set are created for corresponding purposes. To assess the performance of the trained model a metrics computation function is created. It computes such measures as accuracy, precision, recall, and F1-score estimate from the computations of such predictions and real label predictions. Training configurations are set inside TrainingArguments where parameters like number of epochs (set to 3), batch size for training and for evaluation, number of warm up steps, weight decay and clicking on a directory where the logs will be saved. It is proposed to assess the effectiveness of interventions at the end of each epoch to assess progress. The Hugging Face Trainer API makes it easier to train where the model, training arguments, datasets and the metrics function is called. As we are set to see, this unified approach makes training and evaluation relatively easier. During training phase, the trainer adjusts the positions of the BERT model on training set so that it can improve its real fake news classification. It is common practice to perform validation on the model to make sure it does not over fit. This process gives a fine-tuned BERT-based model to classify all the text from which if we combine with metrics evaluation, it gives a strong framework of text analysis. This aligned with the image analysis model, which enables the system to perform the textual and image analyses necessary for identifying fake news.

5.3 Implementation of Image Analysis Model (CNN-Based ResNet)

The image analysis, as a part of the fake news detection system, involves Convolutional Neural Network which uses ResNet-18 as its base model for image analysis, for their classification. The initial step of the implementation involves the initialization of the device; to perform this properly the use of GPU acceleration is recommended. Image transformations are defined to preprocess the dataset, for better data tuning and augmentations that facilitate training. In the case of the training set data augmentation techniques used are rescaling the images to 224 x 224 pixels, randomly flipping the image horizontally for additional variation, converting the image to tensors based on the spatial dimensions of the image and adding/subtracting the mean and standard deviation of the ImageNet database from the pixel intensity values. The same preprocessing is performed for validation and test sets, but the augmentation is not used to keep all sets consistent. The base model used in this study is ResNet-18, and since the network is pre-trained, its ability to extract features based on ImageNet dataset is adopted. In the final layer of the model, fully connected (fc), which was established for thousand classes is changed to make the classes equal to two to classify fake news from real one. This modification returns Gradient boosting by replacing the final layer by a linear layer with two output neurons. The number of input features in this layer is thus made flexible depending on the model structure already existing in order to be compatible. The entire model, right from the updated classification layer, is transferred to the right computational unit – either CPU or GPU. The pre-trained weights advance the convergence and generalization since it is recognized with the identification of common features of images. With the help of the given ResNet-based architecture, as well as the effective preprocessing and data augmentation, the task of analysing the visual information thus presented within the fake news detection process can be performed with reasonable reliability. This CNN model works hand in hand with the text analysis component by emphasizing the visualization component to provide a simplified mode of detecting fake news.

6 Evaluation

6.1 Case Study 1: Fine Tuned BERT Based Transformer Model

The first two sections of Figure 9 are the training and validation loss and the third section is the validation accuracy of the fine-tuning of the BERT model across epochs. First graph titled "Training and Validation Loss" provides a curve of the losses incurred for the training and validation sets. Training and validation losses at the beginning are high making a gradual descent as they epoch progress subsequently, the validation loss makes a gradual ascent. This means the optimization task the model was engaged in as trained was productive, however the increase in the validation loss shows signs that the model may be overfitting it. The second graph titled "Validation Accuracy" presents the ability of the created model in correctly identifying data in validation data set. It begins slightly and then rapidly grows till epoch 1.5 and then begins to fall again.

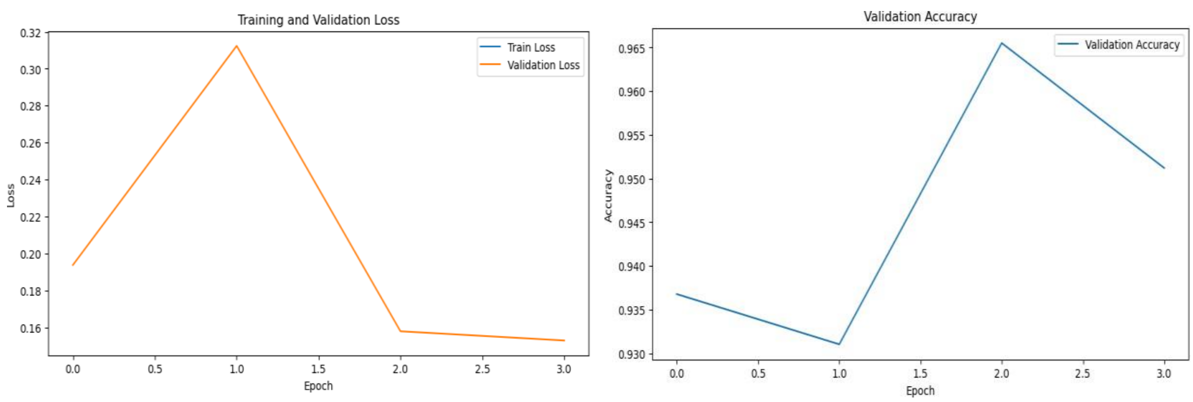


Figure 9: Accuracy and Loss Graph

Table 2: Evaluation Results for Fine-Tuned BERT-Based Transformer Model

Metric	Evaluation Results
Loss	0.1529
Accuracy	95.12%
Precision	95.23%
Recall	95.12%
F1-Score	95.12%
Runtime (seconds)	2.305
Samples/Second	35.575
Steps/Second	4.772
Epoch	3

6.2 Case Study 2: ResNet Model

The training and validation performance of a machine learning model for fake news classification can be seen in two graphs provided in Figure 10. The first one is the plot of the training and validation loss of the algorithm over 10 epochs of training and validation session. The loss values initially are quite high and then oscillate; the training loss typically falls, whereas the validation loss is quite erratic. This pattern indicates how the model is fitting through training data while there are existing some difficulties to generalize on the validation data set. The second graph labelled is “Training and Validation Accuracy”, which shows the accuracy of the model on the training and validation set. Training accuracy initially is low and gradually rises throughout epochs and completes at around 98% in last epoch. The validation accuracy, in contrast, shows more variations and starts from above 96% percent and starts decreasing after that.

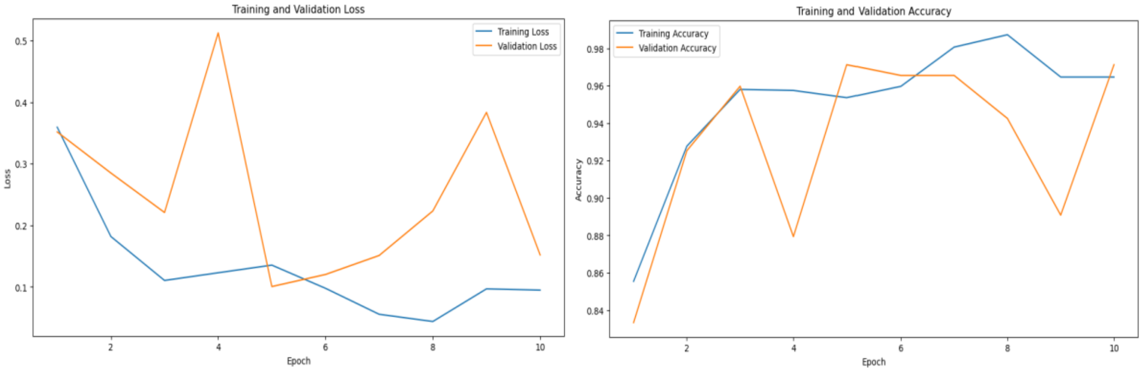


Figure 10: Accuracy and Loss Graph

Table 3: Test Results for CNN-Based ResNet Model

Metric	Test Results
Accuracy	97.56%
Precision	97.67%
Recall	97.56%
F1-Score	97.56%

7 Conclusion and Future Works

This project successfully aligned a multimodal fake news detection system comprising the fine-tuned BERT-uncased transformer for text and CNN based ResNet for the image. OCR integration allowed the text part of images to be easily processed, to extract textual content from the images. The result obtained for both models were satisfactory, the BERT model that we trained scored a test accuracy of 95.12% while the ResNet model tested had a test accuracy of 97.56%. The system is well capable to identify text and image simultaneously which proves the usefulness in handling fake news detection in multimodal environment. The results therefore support the use of the integrated features when classifying text containing a high level of graphics.

Implications:

The presented system's creation has appropriate repercussions for addressing current crucial issues, such as fake news, in digital and social media formats. Proposed as the synergy of text and image analysis, the project targets an important limitation of contemporary approaches to detection based on a single modality only. Media houses, fact checking agencies and social platforms can use the high accuracy possible from the system to reduce the workload on manual detection of fake news. Moreover, it gives an outline for extending multimodal methods to other classification tasks based on the text and images.

Limitations:

Nonetheless, the system has some flaws still; it performs exceptionally well. Firstly, OCR dependence for text extraction is sensitive to the quality and typical image imperfections that may decrease the overall system accuracy. Second, the effective models presented here come at a cost of high training and inference times on the underlying systems, making scalability difficult. Third, the data contains may not sample the entire spectrum of real life fake news, thus the results may not generalize well. Finally, the project lacks an exploration of the temporal and contextual dimension; changes in fake news and its emergence that may limit its use in fast environments.

Future Works:

In light of these, future work might carefully examine the design and execution of the OCR module for enhancing its performance in differentiating noisy or distorted images. Consequently, by including a more extensive and realistic sample of fake news in the dataset, the system's generalization capability can be improved. It might also enhance the accuracy when using temporal and context analysis of how fake news develop to be produced. Furthermore, the current architecture should be engineered by incorporating more lightweight models or optimizing the algorithms in a way that the system becomes resource friendly and can be easily deployed on real world settings. In the last place, such improvements may further increase its functionality in tracking and preventing fake news as it goes viral.

References

1. Abualigah, L., Al-Ajlouni, Y.Y., Daoud, M.S., Altalhi, M. and Migdady, H., 2024. Fake news detection using recurrent neural network based on bidirectional LSTM and GloVe. *Social Network Analysis and Mining*, *14*(1), p.40.
2. Alotaibi, F.L. and Alhammad, M.M., 2022. Using a rule-based model to detect arabic fake news propagation during covid-19. *International Journal of Advanced Computer Science and Applications*, *13*(1).
3. Baarir, N.F. and Djeffal, A., 2021, February. Fake news detection using machine learning. In *2020 2nd International workshop on human-centric smart environments for health and well-being (IHSH)* (pp. 125-130). IEEE.
4. Bento, A.I., Nguyen, T., Wing, C., Lozano-Rojas, F., Ahn, Y.Y. and Simon, K., 2020. Evidence from internet search data shows information-seeking responses to news of local COVID-19 cases. *Proceedings of the National Academy of Sciences*, *117*(21), pp.11220-11222.
5. Irena, B. and Setiawan, E.B., 2020. Fake news (hoax) identification on social media twitter using decision tree c4. 5 method. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *4*(4), pp.711-716.
6. Kalaiselvi, R., 2020. Hot News Detection System Using Rule-Based Technique. *Materials Today: Proceedings*, *24*, pp.2330-2338.
7. Kaliyar, R.K., Goswami, A., Narang, P. and Sinha, S., 2020. FNDNet–a deep convolutional neural network for fake news detection. *Cognitive Systems Research*, *61*, pp.32-44.
8. Krishna, N.L.S.R. and Adimoolam, M., 2022, February. Fake News Detection system using Decision Tree algorithm and compare textual property with Support Vector Machine algorithm. In *2022 International Conference on Business Analytics for Technology and Security (ICBATS)* (pp. 1-6). IEEE.
9. Paschen, J., 2020. Investigating the emotional appeal of fake news using artificial intelligence and human contributions. *Journal of Product & Brand Management*, *29*(2), pp.223-233.
10. Sahoo, S.R. and Gupta, B.B., 2021. Multiple features based approach for automatic fake news detection on social networks using deep learning. *Applied Soft Computing*, *100*, p.106983.
11. Sedik, A., Abohany, A.A., Sallam, K.M., Munasinghe, K. and Medhat, T., 2022. Deep fake news detection system based on concatenated and recurrent modalities. *Expert Systems with Applications*, *208*, p.117953.
12. Shaikh, J. and Patil, R., 2020, December. Fake news detection using machine learning. In *2020 IEEE international symposium on sustainable energy, signal processing and cyber security (iSSSC)* (pp. 1-5). IEEE.