TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Şükrü BIÇAKCI 194283IASM

Fake Content Detection in Social Media Posts

Master's thesis

Supervisor: Sadok Ben Yahia

Imen Ben Sassi

Tallinn 2022

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Şükrü BIÇAKCI 194283IASM

Võltsitud sisu tuvastamine sotsiaalmeedia postitustes

Magistritöö

Juhendaja: Sadok Ben Yahia

Imen Ben Sassi

Tallinn 2022

Author’s declaration of originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature, and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Şükrü BIÇAKCI

03.01.2022

Acknowledgment

Without the scholarships I received from the Tallinn University of Technology - Taltech 2 years ago, I would never have been able to start this project and access the valuable resources that allowed for this breakthrough in Computer and Systems Engineering study.

My sincere gratitude to my academic advisors, Sadok Ben Yahia and Imen Ben Sassi inspired me to push beyond the limits of my knowledge by forcing me to ask tough questions and not stop until I find the answers every day.

I would love to address my family directly with gratitude. Years of unparalleled support and guidance not only helped me define my academic and professional aspirations but ensured that I never lost my faith or determination.

I would like to thank my friend, Emre Arslan, who was incredibly supportive and helpful in finishing this project.

Abstract

In this study, the literature's most recent artificial intelligence models and datasets are listed by conducting detailed research on detecting fake content in social media posts. A mobile and web application has been developed by selecting one of the best of these listed models and datasets. The selected multi-modal yields approximately 82% successful results on the chosen r/Fakeedit dataset. Python - Flask web framework was used for the frontend of the application's technological infrastructure, and React Native framework was used for the backend.

This thesis is written in English and is 60 pages long, including five chapters, 12 figures, and two tables.

Annotatsioon

Võltsitud sisu tuvastamine sotsiaalmeedia postitustes

Selles uuringus on loetletud kõige värskemad tehisintellekti mudelid ja andmestikud kirjanduses, viies läbi üksikasjalikud uuringud võltssisu tuvastamise kohta sotsiaalmeedia postitustes. Mobiili- ja veebirakendus on välja töötatud, valides neist loetletud mudelitest ja andmekogumitest ühe parima. Valitud multimodaal annab valitud r/Fakeediti andmestiku puhul ligikaudu 82% edukaid tulemusi. Rakenduse tehnoloogilise infrastruktuuri esiservas kasutati Python - Flask veebiraamistikku ja taustaprogrammi jaoks React Native raamistikku.

Lõputöö on kirjutatud inglise keeles ning sisaldab teksti 60 leheküljel, 5 peatükki, 12 joonist, 2 tabelit.

List of abbreviations and terms

NLP	Natural Language Processing
DCT	Discrete Cosine Transform
JPEG	Joint Photographer Experts Group
HOG	Histogram of Gauss
EXIF	Exchangeable Image File
CNN	Convolutional Neural Network
RNN	Residual Neural Network
SVM	Support Vector Machine

Table of contents

List of figures

List of tables

1 Introduction

1.1 Motivation

The rapid development in technology has changed people's resources to access information. With the emergence of the Internet, social media has become an essential resource that people worldwide use to access information. Especially in recent years, online social media platforms such as Twitter, Facebook, and Reddit have become popular instead of traditional news sources such as newspapers, television, and radio. The main reason why people use news sources on social media is; their resources are low-cost and easily accessible, and they also enable rapid dissemination of information. For this reason, the number of users who follow the news on social media is increasing day by day. These advantages have created an omnipresent platform for social interaction and information sharing. Social media has facilitated the creation of social groups with millions of members without geographical boundaries. In addition, social media users can share news articles with all group members with the "Share" button on social media. Thus, social media can spread content accessible to millions of people with a single click. Although social media provides many advantages, the quality of news on social media is lower than traditional news platforms. Sometimes, the news contents on social media are changed for different purposes by malicious users. In addition, this type of content is shared and spread by well-intentioned people without being controlled. Therefore, news, content, and comments on social media significantly affect users' opinions. The spread of this type of low-quality news, called *fake news*, affects individuals and societies negatively. Fake news can pose a danger not only to individuals and societies but also to businesses and governments. Therefore, fake news in online social media needs to be identified and detected. Consequently, it is of utmost importance to develop a system that can determine the reliability/credibility of the content in these kinds of posts on social media.

1.2 The objective of the master thesis

According to the motivation of the master thesis, these objectives below will be handled during the thesis studies:

- The most up-to-date artificial intelligence models in the literature will be listed by conducting detailed research on detecting fake content in social media posts.
- Again, by detailed research on the topic, the recent datasets used to detect fake content in social media will be written and compared in detail.
- As a result of the research above, a backend using the selected model will be built. While developing this section, it is intended to assist other software developers and researchers in developing their own tools.
- An interface will be developed where internet users can easily predict the reliability of the social media content they see. It is planned to be a mobile and web application that will harmonize with the backend developed in the previous stage.

2 Literature Review

The most general definition of fake news, Shu et al. [1], has defined a news article intentionally made and is false. They also compared traditional media and social media in terms of fake news. They reported that fake news is usually caused by bots or troll accounts. They defined troll accounts as human-controlled accounts for propaganda purposes and bot accounts as computer-controlled accounts created for propaganda purposes. They stated that these accounts were made quickly and in large numbers. They noted that news shared from more than one bot or troll account simultaneously can be perceived as accurate by ordinary users.

There are two different types of spreading fake news: disinformation and misinformation. If someone is sharing fake news without knowing it is fake, this is misinformation. But if someone is sharing fake news knowing it is fake for some personal, financial, or political, etc. gains, this is disinformation. These people also use fake accounts called “social bots” to share fake news.

Existing studies often use fake news concepts such as disinformation, misinformation, deception, propaganda, satire, rumor, clickbait, and junk news. A glance at these concepts is given below:

- **Misinformation**: It is defined as information that is false or misleading. It can be spread unintentionally due to honest reporting errors or misinterpretations.

- **Disinformation** is harmful content such as illegal speech forms and incitement to violence deliberately created to mislead people.

- **Hoax**: The news is published without being based on true or false news. In general, they involve deception to make people believe an event that did not happen and to fool people.

- **Satirical**: Satirical news is a comedy that mimics news and covers various topics, including social, political, and crime. Written to entertain or criticize readers, these stories can be similarly damaging to hoaxes when shared out of context.

- **Propaganda**: Defined as information that tries to influence target audiences' emotions, views, and actions through deceptive and one-sided messages.

- **Click-bait**: Defined as low-quality journalism that attracts click-through traffic and monetizes through advertising revenue.

- **Junk news**: This content includes various forms of propaganda and ideologically extreme partisan or conspiratorial political news and information. Much of this content is intentionally produced by false reporting. It tries to persuade readers of the moral virtues or failures of institutions, causes, or people and presents the commentary as a news product. This content is produced by organizations that do not employ professional journalists. The content uses conspicuous techniques, many pictures, motion pictures, excessive capitalization, emotionally charged words and images, unsafe generalizations, and other logical errors.

Before detecting fake news, fake news should be thoroughly examined. Then, well-categorized fake news will also be easier to spot. Shu et *al.* [1] analyzed fake news in social media in two different categories. The first of these categories is fake accounts opened for propaganda, and the second is the echo chamber effect, which users define as receiving and sharing news close to their interest, even if it is fake news because they follow users they agree with and trust the news coming from these accounts. They stated that the fake news spread by fake accounts opened for propaganda on social media was spread by social bots, troll accounts, and semi-robot accounts. Many shared news from troll or bot accounts are seen as true for real users and are quickly shared by many real accounts. This increases the credibility of fake news shared from real accounts.

However, social media platforms have improved their algorithms to determine bot activities in recent years, and it has become easier to limit bots' activity. For example, in recent years, Twitter has strengthened its bot detection systems and closed millions of bot accounts to reduce bots' impact and make discussions healthier.

According to one of the surveys of Pew Research Center [2], American people say that the spreading of made-up news is seriously harmful to the country and must be stopped. Nearly 68% of U.S. adults claim that made-up news and fake information impacts Americans’ confidence in government institutions. In addition, almost %54 of people say

it also affects people's confidence in each other. Moreover, most participants think that it will worsen in five years.

U.S. people do not blame journalists for creating made-up news, but they want journalists to fix it. People say that there are two sources for made-up news: political leaders and activist groups. Also, most people think political split causes made-up news to spread.

78% of people say they check the truth of the news themselves, 63% have stopped following just one particular source, 52% have changed social media habits, and 43% reduced overall news intake.

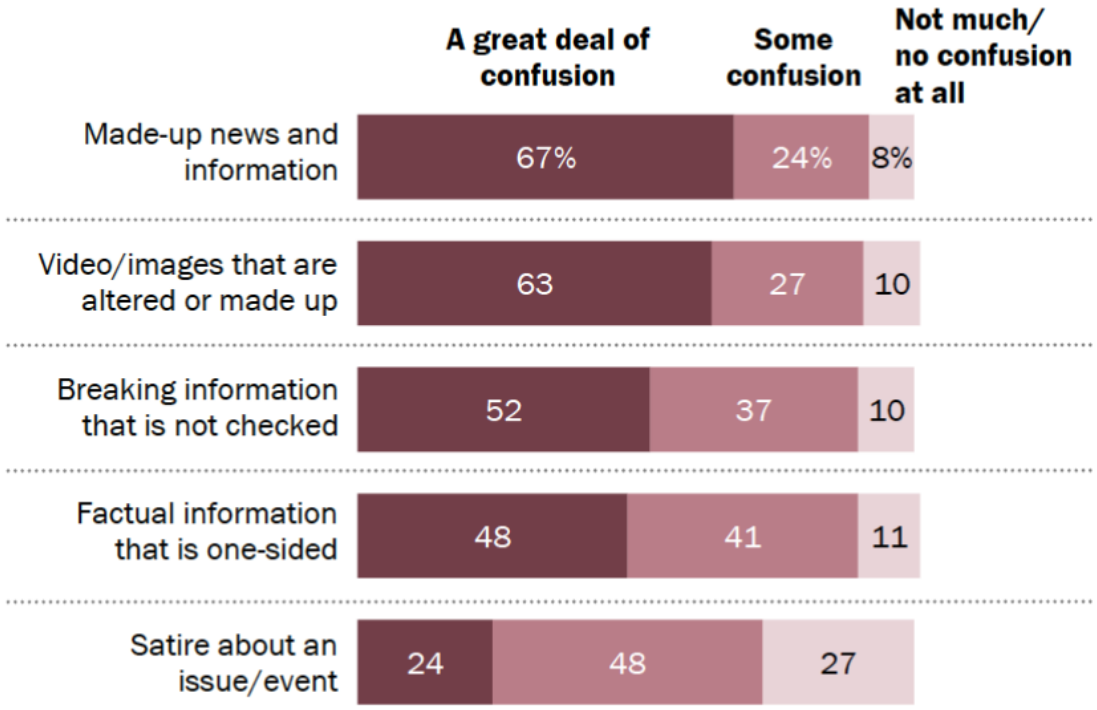

Figure 1.2-1 Two-thirds of Americans think forensic content has a great deal of confusion around facts about current issues.

According to the thoughts of Americans, there are two major topics for made-up news: politics and elections (73%) and entertainment and celebrities (61%). People say that made-up news is often made about national rather than local issues. 58% say so much made-up news is made on national matters, 18% say the same about local issues. Nearly half of Americans think that made-up news and information is a massive problem of informing people. In addition, 51% of people believe the ability of the public to distinguish between truth and thoughts is a huge problem.

Fake news is nothing new, but with the power of social media platforms, fake news is spreading as fast as possible. This is because the journalists are licensed on the official news platform and publish the news with their names. Nevertheless, people can just post images and texts anonymously on social media platforms.

In 2020 Collin, *et al.* surveyed the approaches to fighting against fake news on social media [3]. They described fake news in different types: Clickbait, propaganda, satire parody, hoaxes, etc. Also, they did mention some detection models for fake news: Professional's fact-checker approach, crowdsourced approach, machine learning approach, natural language processing technique (NLP), hybrid technique, expert-crowdsource approach, human-machine approach, graph-based method, deep learning approach, and recommendation system approach. However, most of these approaches have a lot of limitations, huge data sets since they are working manual. So, automatic systems like machine learning, deep learning, etc., are required.

As a result of the developments in image and video processing in recent years, successful image and video editing programs have been used on multimedia. By utilizing these editing programs, even people who do not have basic image or video processing knowledge can easily make the changes they want on photos, pictures, and videos. Employing these editing programs, while end-users make changes on pictures and videos for entertainment purposes in daily life, professional media organizations (magazines, newspapers, news sites, televisions) use them to make images better or more enjoyable. On the other hand, malicious people commit different crimes using such programs on images. This negative situation: In computer science, information security, image and video processing, signal processing has created a new field of study that makes a common working area.

Although the problem of detecting fake news is a new area of research in social media, it has received considerable attention in recent years. Researchers from different perspectives have addressed the issue. According to the authors in [4], The problem is gathered under three main approaches. These are forensic, single-modal, and multi-modal approaches.

2.1 Forensic Approaches

News, information, or any kind of media shared online plays a very integral part in the lives of individuals or societies as a whole. However, widely available tools to easily edit these contents are growing concerns about attaining trustworthy media. Thus, preventing the spread of such content has become the number one priority in recent years.

Vivid and easy-to-understand images often get more attention than text expressions, and they spread quickly among people on social media platforms. For this reason, for the news to reach more readers, visual expressions are added to the news articles, and their content is made remarkable. Visual elements are essential in spreading fake news among social media users. Fake news creators often add fake images and videos to their news content to attract readers. However, such images and videos spread on social media present false information to readers and mislead them. Different characteristics of fake news can be found with this type of visual cues. Other methods are proposed based on the image to detect fake news in social media in the literature.

Image forensic methods detect forgery by using the characteristic features of the image (angle, light, etc.) and statistical information. The most common image forgery methods are copy-move and image-splicing. The copy-move forgery carried out the move from one region to another on the same image. In image-splicing forgery, two different pictures are combined into a single image. Unfortunately, most of the time, it is not possible to determine the changes made on the image with a naked eye since image editing programs perform these processes in a very professional way.

In this survey, Pasquini *et al*. [5] aimed to report and explain the results of multimedia forensic analysis of digital content shared online. Also, to put forward the ongoing issues and challenges needing to be addressed. The survey's review consists of forensic analysis, platform provenance analysis, and multimodal analysis.

Forensic analysis focuses on acquiring the source of the shared media by identifying the source camera or through forgery detection. Unfortunately, the forensic analysis presents due to the applied post-processing operation. As a result, current forensic analysis methods suffer from performance deterioration. This void needs to be filled in to have credible source identifications and forgery detection in future works.

Platform perseverance analysis focuses on reconstructing the sharing history of a certain piece, as it can be uploaded many times on different platforms or from various sources to trace and identify sources. This work uncovers the influence of sharing operations and the odds of predominantly distinguishing its traces and concluding information from it.

Multimedia verification analysis processes various information cues (visual, textual, propagation, and user cues) together and feeds them to a machine learning classifier or decision merger system. This work brings forth the shortage of situated datasets due to the hardships faced in collecting realistic data and the interpretability of detection tools.

On social media, people spread fake news, but they also spread fake images. This is because the visual attachments of news play a significant role in drawing attention. Therefore, these fake images could reach thousands, create harmful situations for users and the public, and cause provocation. Thus, detecting fake news, fake images became a must [5].

In the copy stone counterfeit detection method proposed by Kumar *et al.* in 2015 [6], a gray level image is obtained by combining image color channels. After dividing the gray-level image into overlapping blocks, feature vectors are extracted for each block using DCT. The extracted feature vector coefficients are converted to binary DCT coefficients. The resulting feature matrix is sorted on a row basis, and similar blocks are matched. Certain morphological operations are applied to reduce mismatches. The proposed method is robust against JPEG compressions, blurring, and small-scale rotations.

Lee *et al.* [7] obtained the feature vector by applying HOG (Histogram of Gradients) to each image block after dividing the gray-level image into overlapping blocks. Then, similar blocks are sorted sequentially after sorting the feature matrix consisting of the feature vectors they obtained. To eliminate mismatches, in the final processing step, by hovering over the image with a $16x16$ block size window, matched points less than a certain threshold within the window are assumed as false positives and deleted.

Huh *et al.* [8] proposed a learning algorithm to detect changes/distortions in the trained visual image using a large dataset of actual photographs. The algorithm uses the automatically saved photo EXIF metadata as a control signal to train a model to determine whether an image is consistent. That is, its content can be reproduced with a single viewing line. The authors stated that the proposed method achieved the most advanced

performance for forensic comparisons even though it did not see any altered appearances in training.

2.2 Single-Modal Approaches

With the development of technology, especially computing, social media platforms have become the leading news source. However, these platforms are also available to spread fake news since they are easy to reach and publish.

Deep learning is the application of artificial neural networks to learning tasks using networks consisting of multiple layers. Deep learning methods, which is a new field of machine learning, have attracted attention in many different areas such as image classification, social network analysis, text mining, computer vision, speech analysis, and natural language processing. Moreover, deep learning methods are effectively used in complex and dynamic social media problems. For example, many studies in the literature show deep learning methods are used to detect fake news in social media[9].

In this context, deep learning approaches became one of the best options to detect fake news. In early 2020, Kaliyar *et al.* [10] introduced a new approach for detecting fake news: FakeBERT. This approach works by combining different parallel blocks of the single-layer CNNs with BERT. Generally, researchers are looking at a text sequence in one way, but Kaliyar *et al*. did other in this paper. Their research proposes a BERT-based deep learning approach with different parallel blocks of the single-layer CNN. They say FakeBERT is working with an accuracy of 98.90% better than the existing models. Also, with the model they tried with GloVe, they have achieved an accuracy of 89.97%. So, BERT is more powerful than GloVe.

Microblogs like Twitter or Chinese Weibo offer journalists and their users' content. Not only do users circulate these contents, but they also generate spontaneous news on these platforms. The need to verify this news emerges as a natural consequence.

Image features are essential parts of the microblogs as the text features are limited to specific lengths and are trendy and attract more attention. Therefore, images play an important role in news verification.

In 2017, Jin et *al.* [11] focused on image features for fake news detection as a first attempt at studying image features to validate news on microblogs. They propose a set of visual features and statistical features. Visual features consist of five features: visual clarity score, visual coherence score, visual similarity distribution hologram, visual diversity score, and visual clustering score. These features identify image dispersion characteristics from distinctive patterns of images. Statistical features sum up image statistics and obtain image dispersion patterns quantitatively through seven features. 50,287 tweets and 25,953 images in fake and actual news events from Weibo are collected for performance assessment, resulting in 83.6% accuracy in news verification. Using only non-image features, compared to other approaches, boosts accuracy by more than %7.

Manual methods are insufficient to detect fake images since these images are moved, copied, cropped, etc. In 2021, Singh and Sharma proposed a new deep learning method to detect fake images [12]. This method relied on a deep learning convolutional neural network to overcome the aforementioned reasons. CNN analyzes the features of the images and classifies them by using high-pass filters in image processing. The high-pass filters help details to show up clearly. Singh and Sharma used 16 high-pass filters in the first layer.

A convolutional neural network is widespread in image classification. Singh and Sharma used a customized convolutional network with high-pass filters in their study. These filters are also applied with padding to prevent the loss of information.

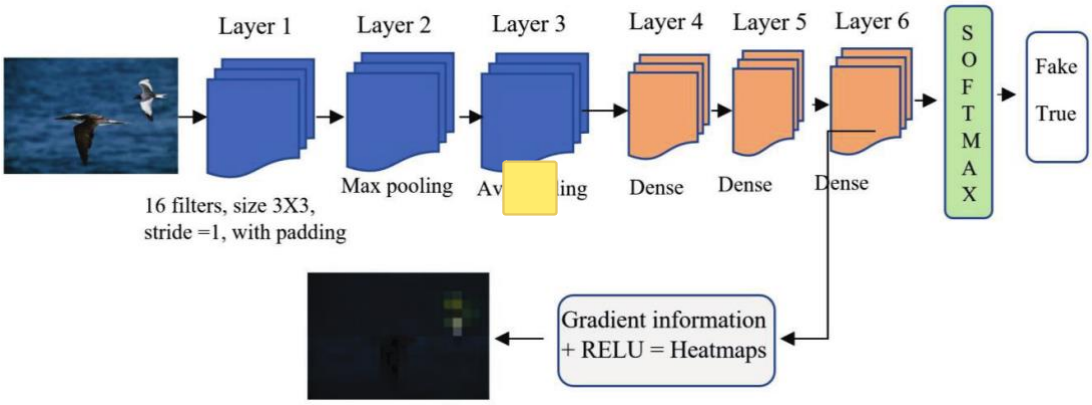

Figure 2.2-1 - Proposed CNN architecture by Singh et al.

This model is compared with CASIA 2.0 dataset. The accuracy of 92% is obtained for CASIA 2.0 and 81.3% over the Twitter dataset that they created. It proves that this model is working well with limitations.

Bhatt *et al.* [13] proposed a new method combining neural, statistical, and extrinsic features to find an effective solution to the fake news detection problem. In the study, neural properties were obtained from the deep repetition model, statistical properties were obtained from the weighted N-gram word bag model, and other external properties were obtained with engineering heuristics. They developed a deep neural network-based fake news detection model by combining the features with the deep neural layer.

In another study, Singhai *et al.* 2017 [14], a deep learning-based automatic fake news detector was proposed through a 3-level hierarchical affinity network. The proposed method has a level for each word, sentence, and headline and creates a news vector representing the news article effectively. The news article is processed in a bottom-up hierarchical fashion to generate the vector. Since the title part of the news contains fewer words, it has fewer distinguishing features for detecting fake news compared to words and sentences.

Fang *et al.* [15] developed a hierarchical neural network model that combines the advantages of convolutional neural networks and the multi-headed self-attention mechanism to detect fake news. First, the multi-headed self-attention mechanism helps the model learn the spatial relationships of words. Later, they combined the attention mechanism with convolutional neural networks. Finally, to prove the validity of their proposed method, they conducted experiments on a publicly available dataset and compared the results with the methods of existing studies.

Qian *et al.* [16] aimed to detect fake news using two-level convolutional neural networks and a user response generator. A bi-level convolutional neural network collects semantic information from news articles displayed at the sentence and word level. The user response generator learns a model that generates user responses to news texts using past user responses to facilitate fake news detection. To test the effectiveness of their proposed method, they used a Weibo public dataset containing accurate news articles and associated user responses.

Girgis *et al.* [17] proposed a new model for detecting fake news in online texts using recurrent neural networks and long short-term memory based on a deep learning perspective. In the proposed model, preprocessing is done on the text, and the data is converted into a format suitable for processing in the model. Then, the extracted features are used as inputs to the repetitive neural networks and long-term memory model.

In a different study, Monti *et al.* [18], a new method for detecting fake news with the geometric deep learning method, a new deep learning class designed to work on structured graph data, is proposed. Geometric deep learning examines disparate data (such as users' activities, social network structure, news dissemination, and content), thus creating a unifying structure for content, social context, and diffusion-based approaches. This model was trained in a supervised manner on real and fake stories spread on Twitter in 2013-2018. The experimental results have shown that high performance is obtained with the proposed method.

2.3 Multi-Modal Approaches

There are two different learning types generally: News content-based learning and social context-based learning. In news content-based approaches, the writing type is essential. The publisher of fake news usually uses a typical writing style to catch the public's attention. Thus, to identify the fake news, fake texts style-based methodologies, catching specific linguistic features are helpful. But it is difficult to determine the fake news with just writing type. In social context-based approaches, the relationship between users and articles is essential. The reaction and behavior of the users to news give a clue about the reliability of the news.

In 2021, Singh and Sharma [4] introduced a new multi-modal approach, which is used to detect fake news. Most of these kinds of models just control the images of the news to detect if it is fake. The news images may be authentic, but also may be beside of point. So, this model checks both images and texts of the news to detect if it is fake or not. For images, the model uses an explicit convolution neural network model EfficientNetB0, and for texts, it uses a sentence converter. There is no need for extra subcomponent support. The convolution neural network model, which is used in this approach, is EfficientNetB0. The new model is way better than others in terms of accuracy in

classification with fewer parameters and lower flops. Also, to control the textual data, this approach uses a bidirectional encoder-based sentence converter named RoBERTa.

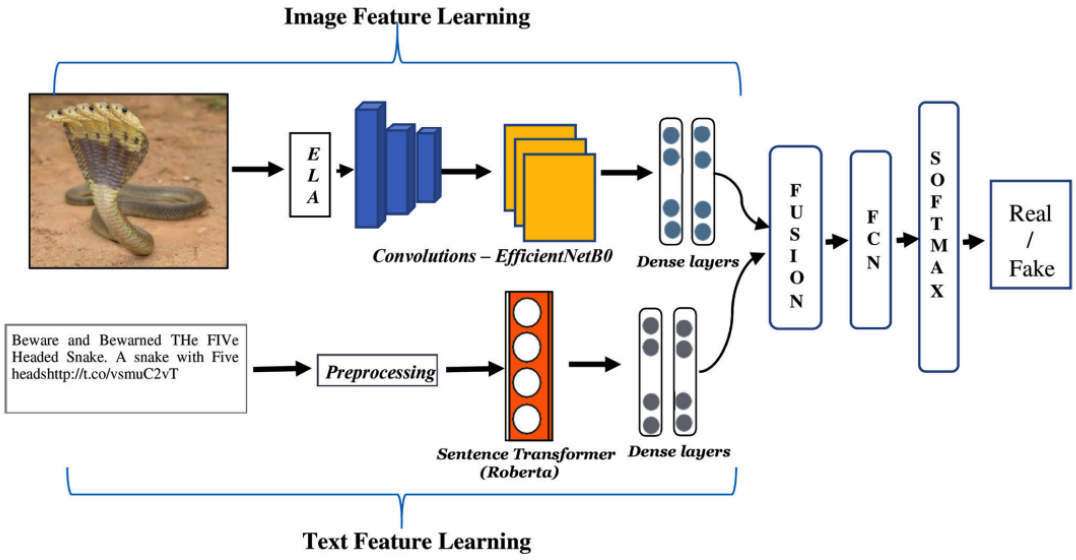

Figure 2.3-1 Predicting image credibility in fake news over social media using a multi-modal approach

There are three different datasets in this model: Casia 2.0 [19], MediaEval [8], and Weibo. All datasets are publicly available. One of them is an image-only dataset, and the others are social media datasets, which involve both images and text. By using all these datasets, Singh and Sharma have tried six different types of EfficientNet, and they noticed that EfficientNetB0 was the best among them. To prove the effectiveness of this model, Singh and Sharma tested the proposed model with these three datasets, and they got an approximative 80%-85% trueness estimation. As a result of all these works and experiences, it can be said that the performance of this proposed model is better than other models in some ways.

For detecting fake news, there are so many different models developed. One of them is EANN. In 2018, Y. Wang et al. proposed a new multi-modal approach named EANN [21]. This model has three major parts: the multi-modal feature extractor, the fake news detector, and the event discriminator. The first part, the multi-modal feature extractor, also has two sub-components: textual feature extractor and visual feature extractor. Each word is treated as a word embedding vector in the textual feature extractor, and these vectors are initialized with the pre-trained word embedding on the given dataset. They use VGG19, a deep learning model with 19 layers in visual feature extractor. The second part, the fake news detector, is built on the multi-modal feature extractor. To detect fake

news, it uses a fully connected layer with SoftMax. And the last part, the event discriminator, is a neural network. This study also used two datasets collected from real social media sites, Twitter and Weibo. With these datasets, to prove the performance of EANN, they tested several methods. EANN has been shown to be the best among them regarding accuracy, precision, and recall.

In the age of social media, fake news or misinformation spreads faster than ever amongst individuals. This makes it a priority to set apart fake news and misinformation. In 2019, Khattar *et al.* introduced a new study detecting fake news [22]. The main target of this study is the detection of news content that is fabricated and can be verified to be false.

Other existing models have a shortcoming in which they do not have any explicit objective function to discover correlations across the modalities. This model, jointly trained with a Fake News Detector, uses Multimodal Variational Autoencoder to detect fake news and uses less information than other baselines. It consists of three main components: encoder, decoder, and fake news detector.

The encoder encodes information from texts and images. Then, the decoder reconstructs learned information back to text and image, and the fake news detector uses the learned shared representation to determine if the news is fake or not.

Two datasets are used to experiment with the model, Twitter, and Weibo, as they are the only available datasets with paired image and textual information. Microblogs have been a popular choice for media consumption in recent years. Among them, Twitter and Chinese Weibo have become essential outlets for this purpose. On these two datasets, results show the MVAE model, on average, outperforms the other methods by margins as large as ∼6% inaccuracy and ∼5% in F1 scores, boosting performance and accuracy.

This paper [23] aims to integrate different content modalities on such microblogs for detecting rumors. Existing methods for automatic rumor detection are based on text and social context. This paper proposes an end-to-end RNN (Recurrent Neural Network) with an attention mechanism, combining text, image, and social context features for rumor detection. Different modalities: tweet text, attached image, and social context provide information. In 2017, Jin *et al.* proposed a novel deep neural network with an attention mechanism (att-RNN) to capture the relations among these. Multimodal att-RNN takes inputting training data, with contents from three different modalities: text, social context,

and image, and outputs a rumor or non-rumor label. Social context, visual, and neural-level attention are critical for achieving the best rumor detection performance by att-RNN. Experiments on Weibo and Twitter datasets show that the proposed att-RNN model can effectively detect rumors based on multimedia content than its existing counterparts based on neural networks.

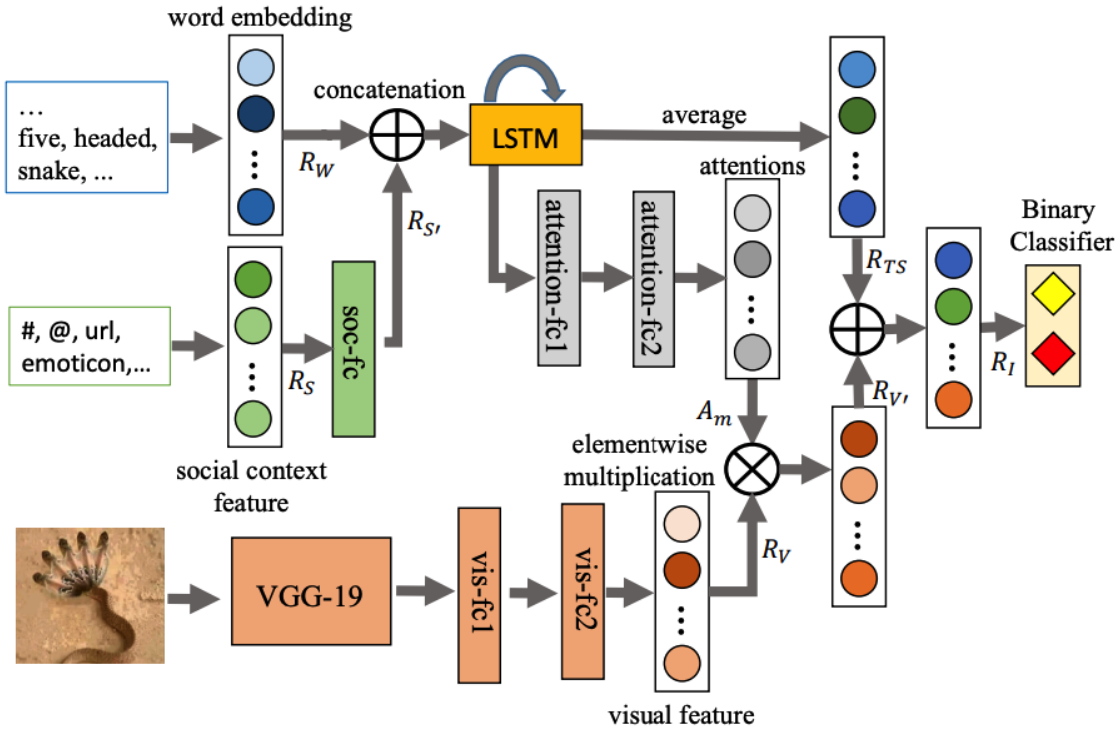

Figure 2.3-2 Multi-modal att-RNN architecture

Social media is now an indispensable part of our life. In an era of countless information pouring online, images, in addition to text, are frequently used to engage more people. This lends itself to the easier dissemination of misinformation. Vishwakarma *et al.* proposed a fake new verification system for fake news on social media platforms [24] with a minimal computational requirement. Moreover, it is an easily implemented and integrated system. This model consists of four units; text extraction from image, entity extractor, scraping the web, and processing the unit. Text extraction from an image performs the extraction of text from the image. The entity extractor extracts entities from the extracted text, and the entities go through various processes of text cleaning. Scraping the web process collects Google search results, labeling reliable or unreliable links based on the calculated value of reality parameter.

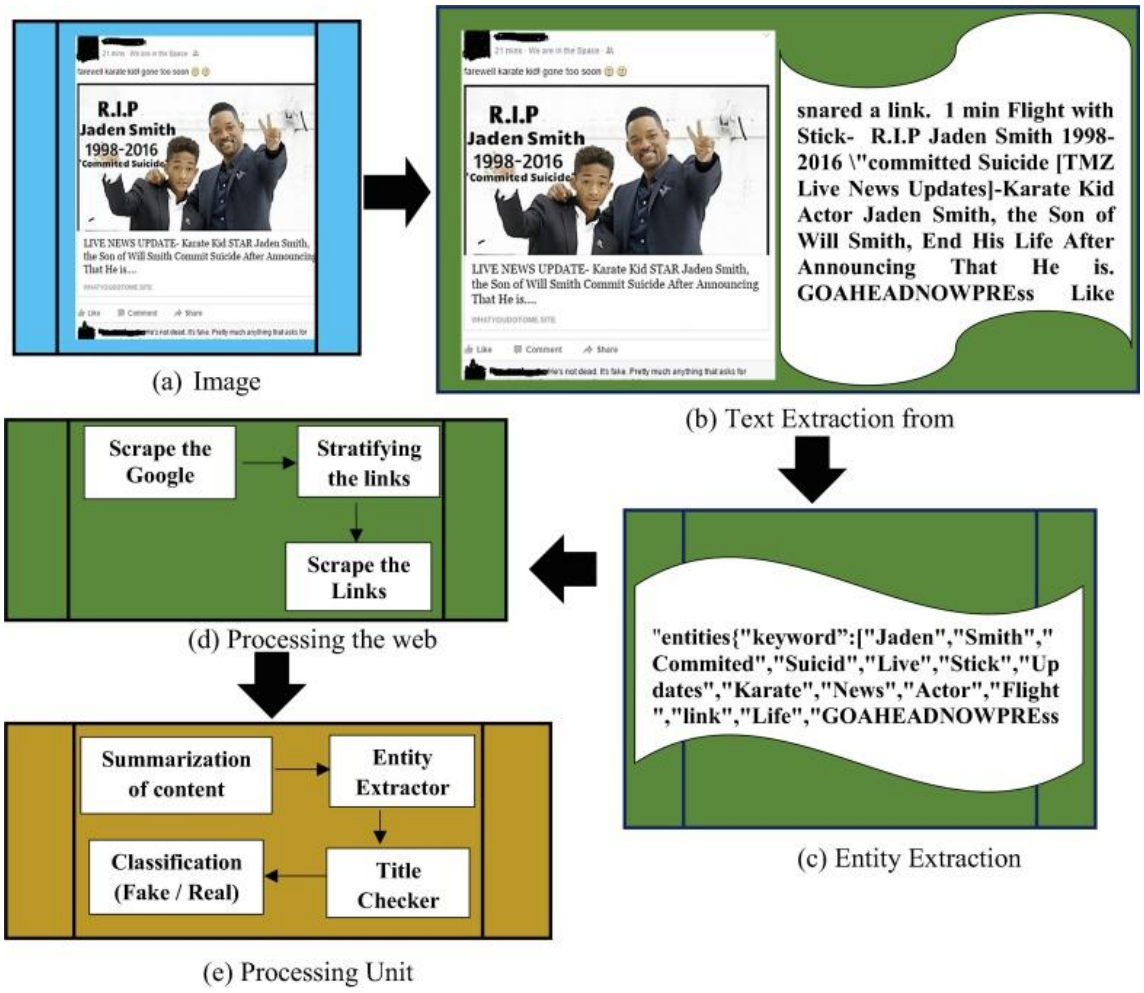

Figure 2.3-3 Detection and veracity analysis of fake news via scrapping and authenticating the web search

Comparing this method [24] with datasets collected from social media platforms like Facebook and Twitter to the other state-of-the-art rumor detection systems, it has been observed that the best accuracy is achieved in detecting fake news via this proposed approach when the Rp (Reality Parameter) value is 40%.

The increasing need for a multimodal fake news detection system grows every day in a social media age. The spread of fake news can have a colossal negative impact on the masses. However, the lack of automated systems and manual methods to prevent the initial spread of misinformation is too slow. Therefore, in 2019 Singhal *et al.* proposed SpotFake [25], a multimodal framework for fake news detection.

SpotFake considers two modalities present in an article, text, and image. It does not consider any other subtasks in the detection process. SpotFake is allocated into three subcomponents: textual feature extractor, visual feature extractor, and multimodal fusion module.

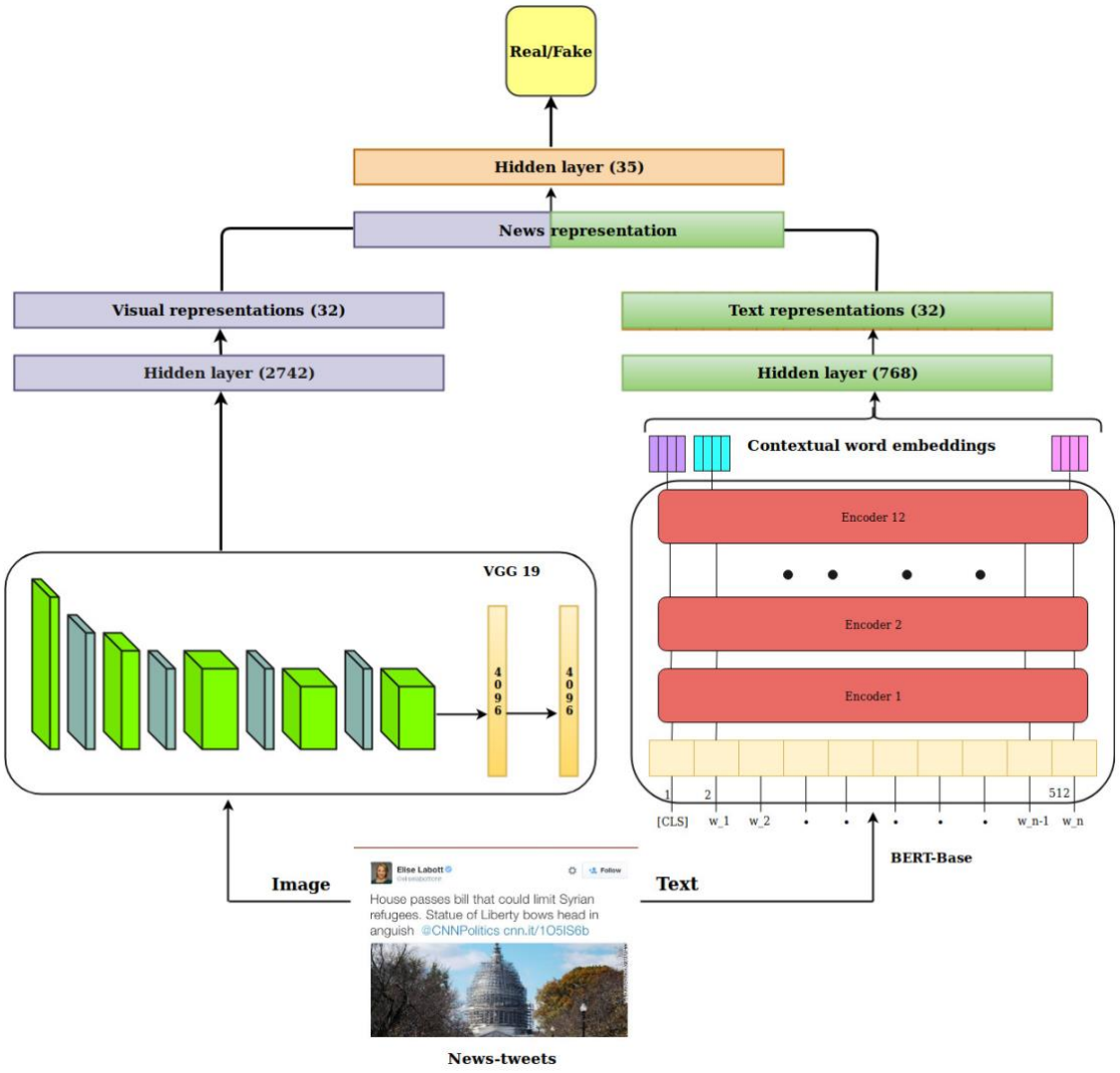

Figure 2.3-4 A schematic diagram of the SpotFake model.

The textual feature is responsible for extracting the contextual text features from the posts using a language model. Visual feature extractor employs VGG-19 to extract visual features. Finally, the multimodal fusion module fuses information obtained from visual and textual feature extractors, creating a news representation. Then the news representation goes through a fully connected neural network for fake news detection.

SpotFake is tested using two datasets, Twitter and Weibo. Against the other state-of-the-art EANN [21] and MVAE [22]configurations, SpotFake sharply outperforms its competitors on both datasets. However, further developments on longer-length articles and more complex fusion techniques can benefit.

The social media growth in recent years lent itself to the fake news problem. Various political agendas and commercial gains feed into this problem. Naturally, fake news detection studies gained traction. However, few studies have been on the impact of user sentiments, particularly user comments. Incorporating user sentiments into a detection process is the main novelty of this study. In 2019, Cui *et al.* presented a Sentiment-Aware Multi-modal Embedding (SAME) [26] which considers both the sentiment and multi-modality. They introduce a deep end-to-end framework to fuse different elements of the news pieces for fake news detection. Over various modalities, a hostile system is added to conserve semantic relevance and representation coherence. User sentiment is verified using statistical analysis and users' emotional contrasts to detect fake news.

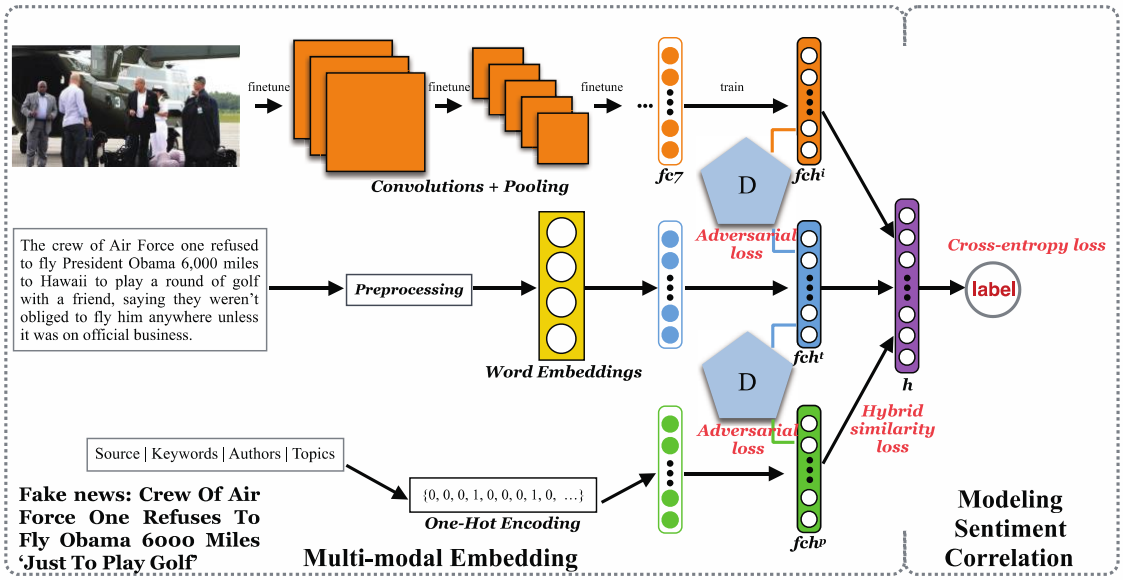

Figure 2.3-5 SAME model

A new fake news detection structure is built with multimodal data, an adversarial system, and user sentiment. Multi-modal feature extractor, adversarial learning, sentiment correlation model, and fake news detector modules are very effective and necessary components of this approach [26].

During the presidential election in 2016, people have noticed that "fake news" is serious than they thought. In the study of Vosoughi *et al.*, they say that fake stories spread faster than real stories. Social media plays a big role in this situation. Generally, fake news detection methods are split up into two: content-based methods and social-context-based

methods. The main difference between them is whether it depends on the social context. It is obvious that having more social context makes fake news detection easier.

In February 2020, Xinyi Zhou *et al.* proposed a similarity-Aware FakE news detection method (SAFE) [27]. This method has three modules: performing multi-modal feature extraction, within-modal fake news prediction, and cross-modal similarity extraction. This method first adopts neural networks to obtain the latent representation of text and visuals according to the similarity. Then, the similarity between the text and visual information of the news and its representations are used to detect whether it is fake or not.

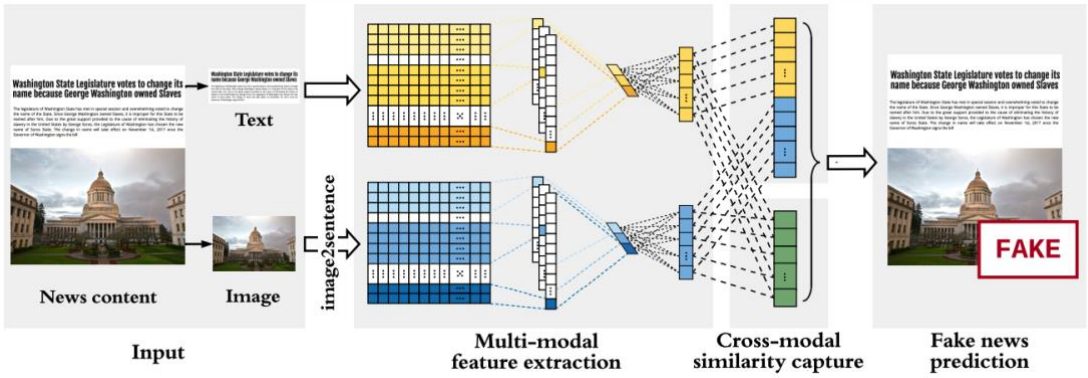

Figure 2.3-6 SAFE: Similarity-Aware Multi-Modal Fake News Detection

In this study, two public benchmark datasets of fake news detection are used. These datasets include news articles of PolitiFact and GossipCop. And the ground truth labels of the news articles, which are partaking in datasets, are provided by experts. In short, this method extracts the news' textual and visual features and analyzes the relationship between them without considering, e.g., network and video information.

Microblogging sites commonly use digital spaces to share news, information, images, or videos. Unfortunately, they are also used to spread fake news through forged images and videos. It has been scientifically proven that images further alter the knowledge our brains consume. Therefore, they are frequently used to disseminate fake news.

In 2021, Singh and Sharma [28] proposed a convolutional neural network model with an attention mechanism to detect fake images on social media. The neural network can detect multiple manipulations on an image by concentrating on the most relevant part and learning from it. High Pass filters are used to reveal hidden features of the forged image. The LIME (Local Interpretable Model-agnostic Explanation) method has been employed

to ensure the model's reliability. LIME brings interpretability to the model and provides an understandable explanation. The model can handle complicated cases involving splicing, face-swapping, text-editing, copy-move, and mirrored images.

To test the model, two datasets were used. CASIA 2.0 dataset [19] achieved an accuracy of 94.4%, and a real-world Twitter dataset achieved 82.3%, proving that a deep learning model is better at detecting forged images on social media platforms. The proposed model can be used automatically worldwide, reducing the manual workforce. However, there are a few shortcomings of the proposed model. Large groups of people or gatherings in images are challenging for the model to predict. The same result is achieved when the images are unclear, blurry, or taken from afar.

Wang [29] proposed a hybrid convolutional neural network model for fake news detection. In this model, two input parameters are taken: the phrase text and the speaker's metadata information. First, the text data is processed by the word embedding layer to obtain continuous low-dimensional representation forms for each word in the sentence. Next, the convolutional and maximum pooling layer processes the output of the layer to produce the feature representation. Similarly, speaker metadata is processed with a convolutional and bidirectional long-short-term memory layer to obtain the final feature representation patterns. Finally, the two representations are combined and feed the end-to-end trained classifier with the other layers. To test their effectiveness, deep learning-based convolutional neural networks were compared with SVM and Logistic Regression algorithms.

2.4 Patents

In addition to the articles mentioned in previous chapters, the following section will share patent submissions about the studies on developing a system to detect fake content in social media posts.

A MACHINE LEARNING APPROACH TO VALIDATE THE AUTHENTICITY OF NEWS USING NATURAL LANGUAGE PROCESSING (AU2021106048A4)[30]

This invention is about a machine learning approach to detect the trueness or fallaciousness of news by using language processing. This approach identifies the news sources and classifies them as fake or not. Fake news detection assignment is separated

as input and output. The input is a one-line statement, short statement, or entire article, and output is mainly classified as news or trust score.

This approach is used for detecting fake news and fake sources by using a combination of machine learning layers. These machine learning techniques are sentiment analysis, text analysis, scoring of articles, etc.

This system detects fake news as the different news sources input the system. These sources are treated with varying processing systems like scoring, semantic and language processing models, consensus-based tracking models, user profile-based news quality detection models.

The system of the invention identifies the essential parameter in news datasets, which can improve the accuracy of fake news detection and investigation model using Natural Language Processing.

FALSE NEWS DETECTION METHOD AND SYSTEM BASED ON MULTI-TASK LEARNING MODEL (CN110188194A)[31]

This invention is about news detection technology based on a multi-task learning model. Since it is difficult to detect fake news just based on news' content, inventors thought that users' auxiliary information like social media activities could be helpful. This method detects news authenticity and subject matter simultaneously using a multi-task learning model. The multi-tasking model consists of an embedding layer, a presentation layer, and a multi-tasking layer. In the embedded layer, the text content, and the context information of the news, which is to be detected as original data, are embedded into a low-dimensional space. In the presentation layer, GRU (Gated Recurrent Unit) layer and a CNN (Convolutional Neural Network) model are used for text feature extraction. Another CNN model is used for the context embedded vector. In the multi-tasking layer, the authenticity of the news is tried to be found, and the topic classification of the news is attempted to be done.

METHOD AND APPARATUS FOR COLLECTING, DETECTING, AND VISUALIZING FAKE NEWS (WO2020061578A1)[32]

Detection of fake news is not easy since the situations like insufficient datasets, the dynamic nature of fake news, etc. A fake news detection system could work better with better and more extensive datasets and a model that classifies the news. Some systems have recently detected fake news, but almost all of them require human intervention in the fake news detection process. This invention has a system for online news collection, detection, and visualization of fake news. There are three steps in this patent application: Fake news data collection, fake news detection, and fake news visualization.

- In the first step, the invention collects verified fake news and true news from fact-checking websites. Then, using the APIs of social media platforms, it searches and gathers the social media posts like tweets, etc. Also, it searches and collects the interactions of social media users like a retweet, repost, like, etc.

- In the second step, the invention detects fake news by using Social Article Fusion (SAF) model. This embodiment of the invention uses the linguistic features of news content and features of social media context to classify fake news. And a second embodiment of the invention benefits the relationship between publishers, news, and social media engagements. And third embodiment examines the relationship between profiles of social media users and fake news.

- In the third and last step, embodiments procure a web-based visualization to analyze the collected dataset. Word cloud visualization allows for seeing fake news and real news topics.

2.5 Datasets

In the study [29], in which Twitter's 12-year data was collected and the tweets examined by six independent verification sites since 2006, it was determined that the news containing 126 thousand false information was shared 4.5 million times by 3 million people. False news routinely reaches more than 10 thousand people.

In 2013, Jing Dong et al. introduced a new dataset for image tampering detection. The benchmark database name is CASIA [19]. This dataset has two different versions. These are CASIA ITDE V1.0 and CASIA ITDE V2.0. Version 1.0 is smaller and contains 1,725 color images in total. 921 of these pictures were tampered with, and the remaining 800 were left untampered. Images are 384 x 256 pixels size and JPEG formatted. The larger Version 2.0 comes with 12,323 color images. While the tampered set consists of 5,123 colorful images, the untampered set contains 7,200 colour images. The image dimensions are not fixed compared to the first version in this version. Moreover, the image has been manipulated harder.

It is known that fake news is bad not only for individuals but also for society. Reliable news is essential, especially when getting information about public safety. But with the increasing number of society-based news sources such as Twitter, Instagram, etc., it became harder to reach reliable and real news. For this reason, automatic verification and cross-checking tools became more required. At MediaEval 2015, by Christina Boididou *et al.*, a new study is introduced [20], which is developed to form a basis for future generation tools of the process of verification.

In this study[20], a new task is mentioned. In the latter, participants were given a list of tweets which included an image or a video of a popular event that gets attention. Then, the participants were asked to guess whether these tweets were fake or real. A tweet, which includes a photo or video that does not represent the event, is fake. On the other hand, a tweet is accurate, consisting of a photo or video from the event. Also, participants were allowed to regard a tweet as unknown.

There are two different datasets in this study. The first is the "development dataset," also named "devset." It involves tweets that are related to 11 different events. As a result, there are 176 cases of real and 185 instances of misused images. The second is "test dataset," also named "testset." This one is used for evaluation. It involves 17 cases of real images, 33 of misused images, and 2 cases of misused videos. In this task, the sought-after goal was to find an automatic method to distinguish between two types of multimedia in tweets: reflecting reality or spreading fake impressions.

Widely used and available social media platforms leave an open door for spreading fake news. Often, this fake news is about important political figures or famous individuals. It's

important to prevent such news to preserve these people's reputations. Fake news detection models started being developed for such purposes. However, it has proved challenging because of the shortage of labeled data and the purposefully misleading manner of writing. In their paper, Jindal, Saad *et al*. presented two benchmark multimodal datasets with image and text. At first, the training dataset, named NewsBag, is created with 215,000 news articles (200,000 real and 15,000 fake news) extracted from *The Wall Street* for the real news and *The Onion* for the fake news [30]. However, this dataset is imbalanced. Therefore, they created a more extensive and appropriately balanced training dataset called NewsBag++, with 589,000 news articles (200,000 real and 389,000 fake news) using a data enhancement algorithm. NewsBag++ is primarily used for producing fake articles to further train modals. NewsBag Test containing completely new 11,000 real and 18,000 fake news is created for the sole purpose of testing the datasets while the NewsBag or NewsBag++ trained the models. The main weakness of this dataset is not considering any social context like sharing trends or user comments. Training single modality or multimodal models with this dataset reveals the difficulty of detecting fake news by showing the weak generalization abilities of these models. It also shows there is room for improvement on widening the modality sets in fake news detection datasets with audio, video, or social contexts.

With the growth of social media, the spreading of news is increased, but the spreading of fake news is also increased. Therefore, the development of applications that detect fake news has become inevitable. But, unfortunately, the inadequacy of datasets is one of the biggest problems against this development.

In 2020, Kai Shu *et al.* introduced a new data repository named FakeNewsNet [29]. The FakeNewsNet contains two datasets with various features in news content, social context, and spatiotemporal information. Kai Shu *et al.* followed a process to collect news to create and contribute to FakeNewsNet;

- **News context**: To collect ground truth labels, they use PolitiFact and GossipCop. In Politifact, the experts provide evaluation results to decide the news' article is fake or not. In GossipCop, there is a rating for every article. Their study accepts GossipCop's news as fake, rated less than 5.

- **Social context**: They are also analyzing the tweets that they collect. The social impact of these tweets is important to decide whether the tweet is fake or not. The likes, retweets, replies are considered. Also, the social profiles of the users are considered too.

- **Spatiotemporal Information**: The spatiotemporal information consists of spatial and temporal information. They collect the user profiles' location for spatial information. They examine the serial change of the user profiles, spreading fake news, for temporal information. FakeNewsNet is highly extensive and collected from different sources and helps the fake detection approach with its datasets.

With the rising of social media applications and other online sources, which are doing inadequate fact-checking, the creation and spreading of fake news increased. According to a Pew Research Center, nearly half of American people consider fake news a critical problem and worse than violet crime. For this reason, people are trying to detect fake news and doing much research for it.

To create a fake news detection model, sizable and diverse training data is a must. There are some published datasets, but these datasets also have limitations like size, modality, etc. Therefore, in 2020 Kai Nakamura et al. proposed a new dataset: Fakeddit. r/Fakeddit [31] is a novel multimodal fake news detection dataset created with 1 million samples with 2-way, 3-way, and 6-way classification labels. For example, 6-way classification consists following labels; True, Satire/Parody, Misleading content, Manipulated Content, False Connection, and Imposter content.

The researchers defend that this dataset will expand fake news detection into the multimodal space and help the fake news detection systems be more generalized, fine-grained, and well developed. With neural network architectures, which integrate the image and text data, they evaluate the dataset through text, image, and text+image modes. Also, Reddit, a social news and discussion website, is used as a source for this dataset.

The table below shows all researched datasets by their content. Datasets contain images, text, or other types of helpful content to use in multi-modal approaches. Those other types of content are post comments, user profile info, post-like numbers, etc.

Table 2.5-1 – Table of Datasets

Name	Content		
	Image	Text	Other
MediaEval	X		
Newsbag		X	
FakeNewsNet	X	X	X
CASIA	X		
r/Fakeeddit	X	X	X

3 METHOD

In this chapter, two branches of the developed method are discussed. These are the neural network model and mobile and web application.

3.1 NEURAL NETWORK MODEL

This model is the pre-trained model created by [36]. The model was trained with the r/Fakeeddit dataset and provided 82 percent accuracy in the 6-way classification of the r/Fakeedit dataset. Therefore, we selected the text+image multi-model suitable for my application and used it as its prediction engine.

3.1.1 Textual Modal

DistilBERT pre-trained model [37]was used for the selected method's textual feature detector. Before explaining this model, it is necessary to explain the BERT model[38].

In 2018, Google announced Bidirectional Encoder Representations from Transformers BERT. As the name suggests, unlike other models, it evaluates the sentence from left to right and left. In this way, it plans to understand the meaning and relationships between words better, and the results pay off.

Using the dataset BookCorpus with 800M vocabulary and Wikipedia with 2.5B vocabulary, two basic models called *bert_large* and *bert_base* were presented. Bert_large was trained with 16 TPUs, and bert_base was trained with 4 TPUs for 4 days.

Transfer Learning from large-scale pre-trained models is becoming increasingly common in Natural Language Processing (NLP), yet running these huge models on edge and/or with restricted CPU training or inference budgets remains difficult. Therefore, the authors created a new model to solve these challenges by distilling the BERT model. The name of this new model is DistilBERT. It is a distilled version of the BERT model. In Figure 3.1-1, How DistilBERT evolved from BERT is shown.

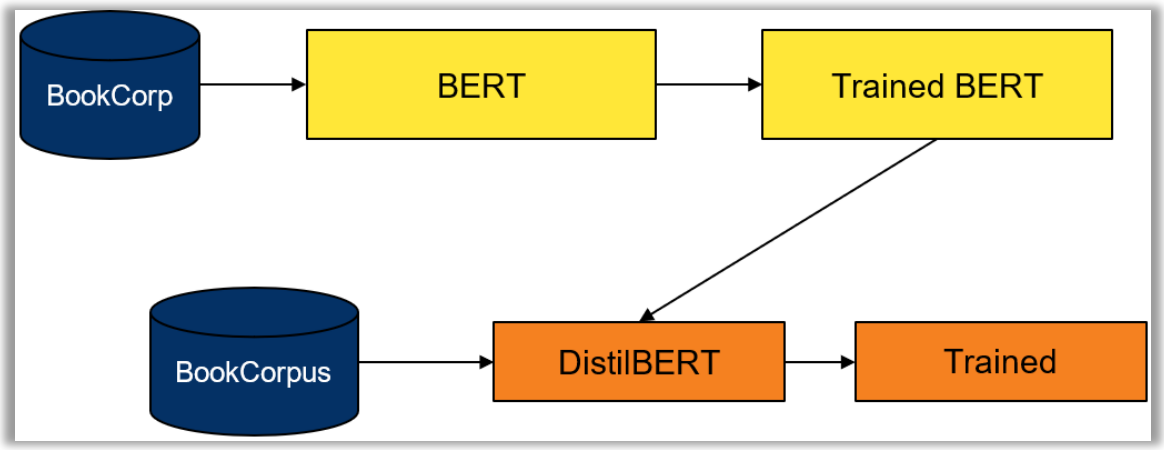

Figure 3.1-1 DistilBERT distilation schema

Overall, the distilled model, DistilBERT, has about half the total number of parameters of the BERT basis. It is 60% faster than the BERT base model and retains 95% of BERT's performances on the ELMo + BiLSTMs language comprehension benchmark.

Table 3.1-1 BERT vs. DistilBERT

	# of parameters (millions)	Inference Time (s)
ELMo + BiLSTMs	180	895
BERT base	110	668
DistilBERT	66	410

3.1.2 Image Modal

The ResNet34 model trained with ImageNet [39] 2012 dataset was used for image features in the chosen method.

ResNet [40] is a neural network structure proposed by He Kaiming, Sun Jian, and others from Microsoft Research Asia in 2015 and was the winner in the ILSVRC-2015 classification task. It also took first place in ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation tasks.

To create a Residual Network, we combine Residual multiple blocks. With this concept, the Researchers created multiple variants of the Residual Network with different layer numbers such as ResNet34 and ResNet50. For example, the following architecture for ResNet34 34-layer residual with skip connection can be seen in the figure below.

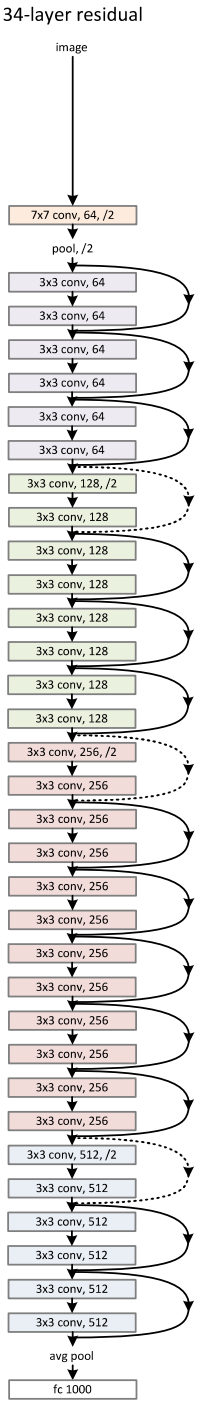

Figure 3.1-2 ResNet34 architecture

3.2 MOBILE AND WEB APPLICATION

This chapter shared the technologies we use for the backend and frontend and the development phase while developing your mobile and web application.

3.2.1 Backend – Flask

A web framework is a library or collection of packages or modules that can save lives when building scalable, reliable, and sustainable web applications. Frameworks make it easy to reuse code in common operations by avoiding code clutter.

Flask is a Python framework [41]. As it is known, Python is a life-saving language to do something quickly and to reveal specific results by saving time. In web services, the flask framework of Python can be used to get fast results. Flask is a framework that can be learned quickly and has a high performance when looking at its benchmarks.

Flask Pros:

- Extremely flexible
- Easy to learn and use
- Redirect URLs easily

You want to enter a website, and you create a request by entering this site address in the URL of your search engine. The request you create reaches the server by getting the static IP of the site from the DNS servers. The server also sends a response to this request. We receive this response and display it in our search engine in HTML form. Our back-end application will work similarly to this working principle.

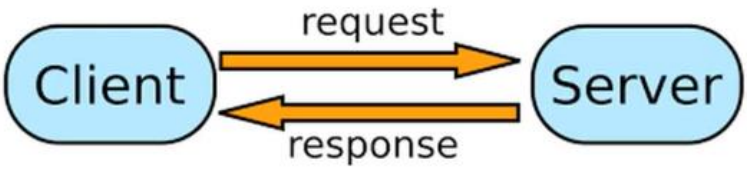

Figure 3.2-1 Working schema of a web server

To install Flask, you must first open your terminal and type the code block below into your terminal.

$$pip\ install\ Flask$$

Pip is a package manager that comes with Python. It is a package management tool that allows you to manage the libraries, framework, etc., dependencies that do not come with Python.

We can import the Flask web framework to our Python application as follows. In this way, we can use the built-in functions in the framework and create our own application.

$$from\ flask\ import\ Flask$$

Then we will use the features of the flask library in our project by creating an app object from the Flask class.

$$app = Flask(\_\_name\_\_)$$

Our Python application has three different request/response functions. These are:

1. def $initApp( )$:

   This section runs after the request is sent when our front-end application is started. Here, our neural network model is created as an object. Then, the files that need to be downloaded in the background are downloaded and uploaded. When

everything is completed successfully, "start" is sent to our front-end application as a response, and the main application screen is shown to the user.

2. def $download()$:

   In this section, the picture selected by the user in our front-end application is taken with a "POST" request and saved to the server locally. This picture will then be used to be predicted by our neural network model.

3. def $predict()$:

In this section, the text written by the user to the textbox in the front-end application is retrieved. Then, the image currently downloaded by our last *download()* function is read. Then our neural network object, which is already created by *initApp(),* is given this text and image and made to predict them. And we send this prediction to our front-end application as a response, and it is shown to the user as an alert.

3.2.2 Frontend – React Native

We can briefly say that React Native is a framework produced by Facebook that enables cross-platform mobile application development [42].

React Native is a mobile application development framework (software) that enables the development of multiplatform Android and iOS applications using native user interface (UI) components. It is based on the JavaScript Core runtime and Babel transformers. React Native; Supports new JavaScript (ES6+) features such as arrow functions, async/await.

This framework for mobile application development started in the summer of 2013 in the Facebook Hackathon Project. It was first introduced at the Reactjs Conference in January 2015, and in March 2015, Facebook made React Native open and available on GitHub.

Since then, it has been adopted by most developers and organizations due to its ability to create successful user interfaces with native applications. You can observe the uptrend of React Native in the chart below. Only 1.5 years after its release, it has surpassed Android and iOS development.

Therefore, it should be no surprise that most applications we use today (UberEats, Facebook, Instagram, etc.) have a logic built mainly using JavaScript rather than Java / Kotlin or Objective-C / Swift.

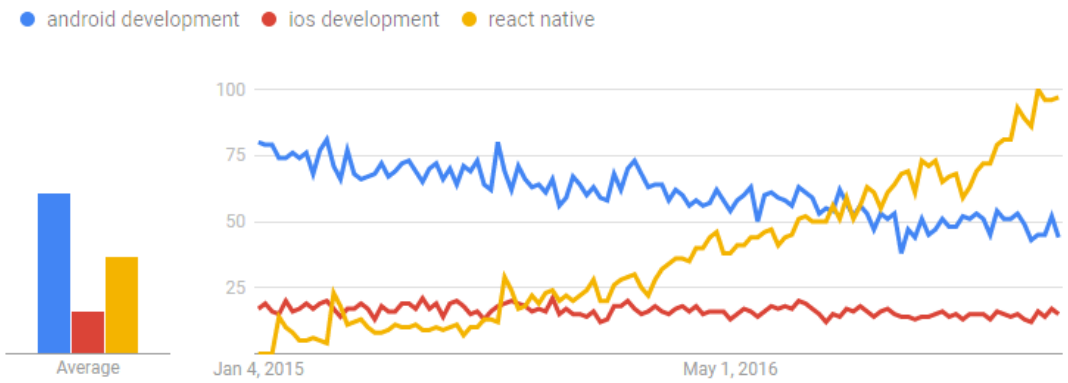

Figure 3.2-2 Compared with iOS and Android Development Trends

In general, React Native advantages can be listed as follows:

- We can release applications for both Android and iOS platforms through a single platform with a single language and multiple platforms. If we wish, we can create and edit designs independently. This angle is a big plus.

- Community support is very high. After Facebook released React Native, it was loved by the developers. Since it is a new framework, it had many errors and shortcomings, and these were fixed and improved over time, thanks to the developers. There are still more recent packages, new versions being released, and Facebook responds to these requests with each new update.

- One of the most important disadvantages of hybrid mobile applications is the slow loading speed. The loading time of a mobile application developed with React Native technology is much shorter. This technology gives functional results in terms of speed in mobile applications.

- When you write a code while developing an application with React Native, you can instantly see the changes you have made in your code with the live reload feature without the need to run it.

Negative Aspects of React Native:

- React Native technology provides great convenience and time savings in terms of code during design. However, it will be challenging to find and debug when some errors occur in an application developed with React Nativeug because you should research the code structure and create the right action plan.

- With React Native, you can write Android applications on Windows PC, no problem, but you cannot write an iOS application. For this, you need a Mac computer or install a Virtual Machine. On the Mac side, there is no such problem. You can develop an Android or iOS application without extra effort.

In the figures below, the usage details of the developed application are shown for iOS and the web platform. Starting from the top left in Figures 3.2-3 and 4; The figure in the upper left shows the screen encountered when the application is opened for the first time. The selected model is expected to be loaded in the backend during this screen. If everything is ok, the user will be directed to the screen shown in the figure in the upper right. On this screen, the user can proceed by pressing the "Predict" button after typing the desired text and selecting the picture (as seen in the figure at the bottom left). After pressing this button, the loaded content is sent to the backend to be predicted, and the result is reflected by the user as an alert (as seen in the figure on the right).

All code not described here can be viewed in Appendix 2 on the GitHub repository.

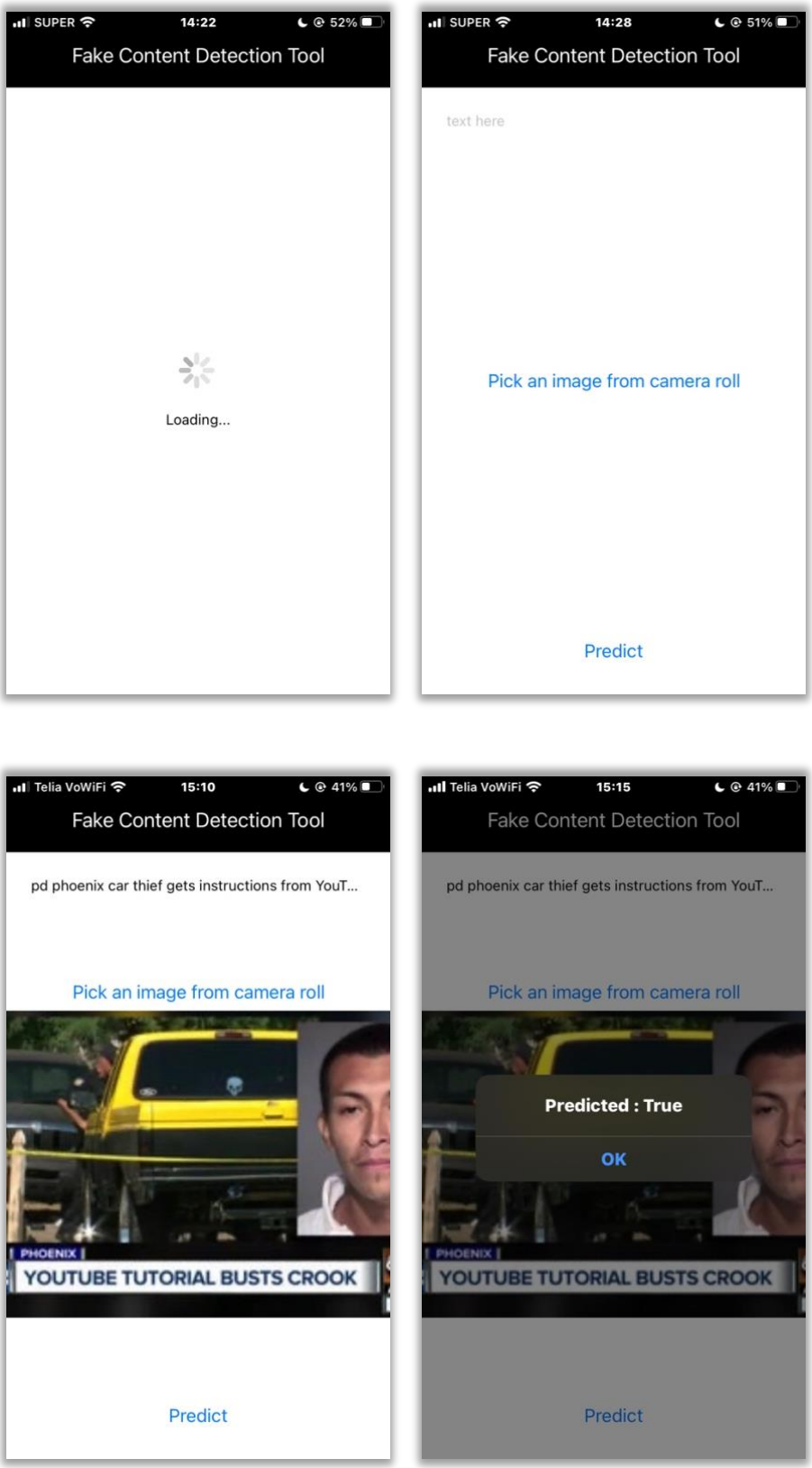

Figure 3.2-3 iOS version of the application

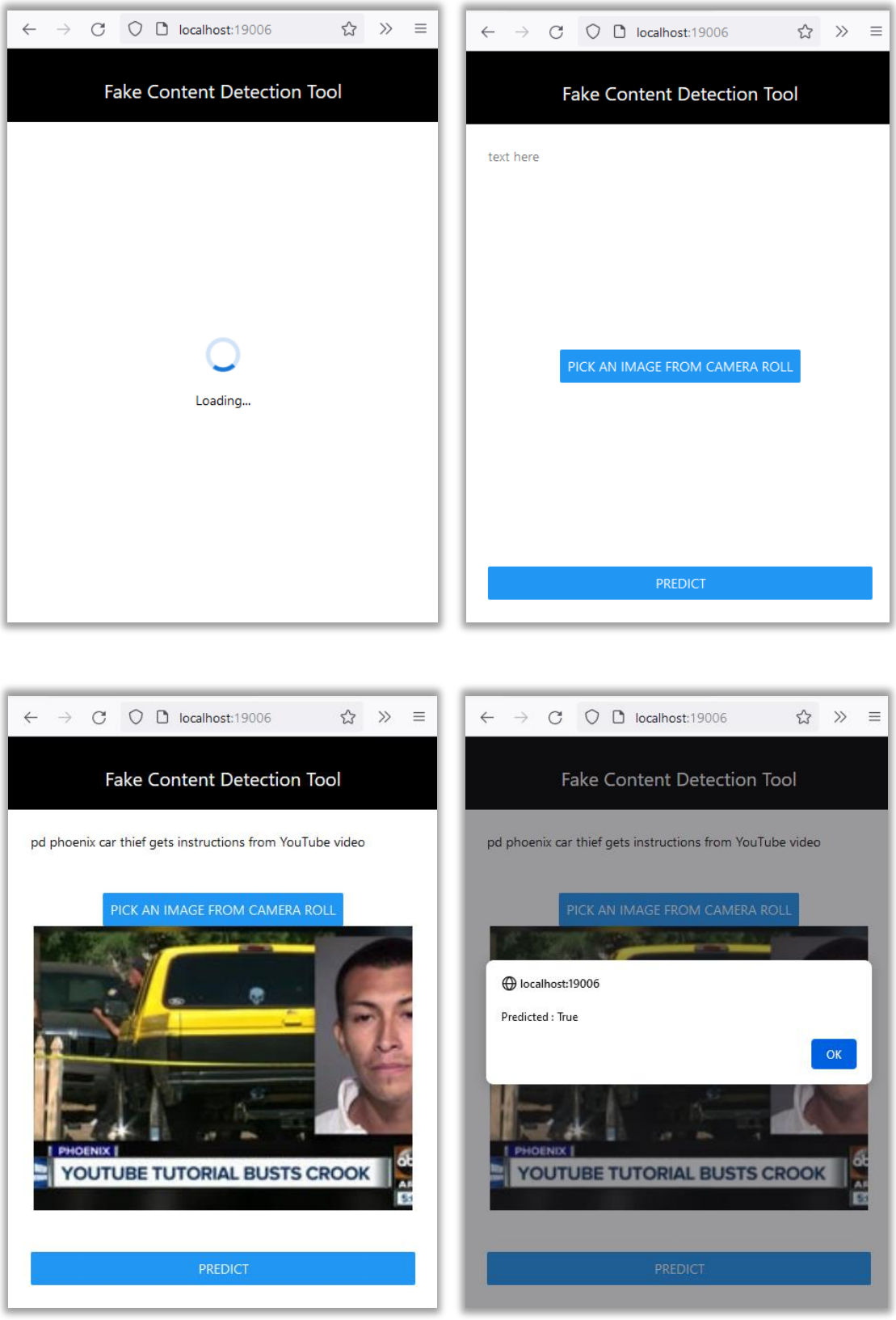

Figure 3.2-4 Web version of the application

4 RESULT

This chapter shares the results of the neural networks model and mobile and web applications, which are two separate parts of our method.

4.1 Neural Network Model Test Results

This section shares the test results of the artificial intelligence model used. The model was trained and tested using the visual and text contents of the r/Fakeeddit dataset. It is seen in Figure 4.1-1 that the best results were obtained in the 14th epoch. This model achieved approximately 82 percent success in the r/Fakeeddit dataset. The author stated only 20 percent of the dataset data was used during the training.

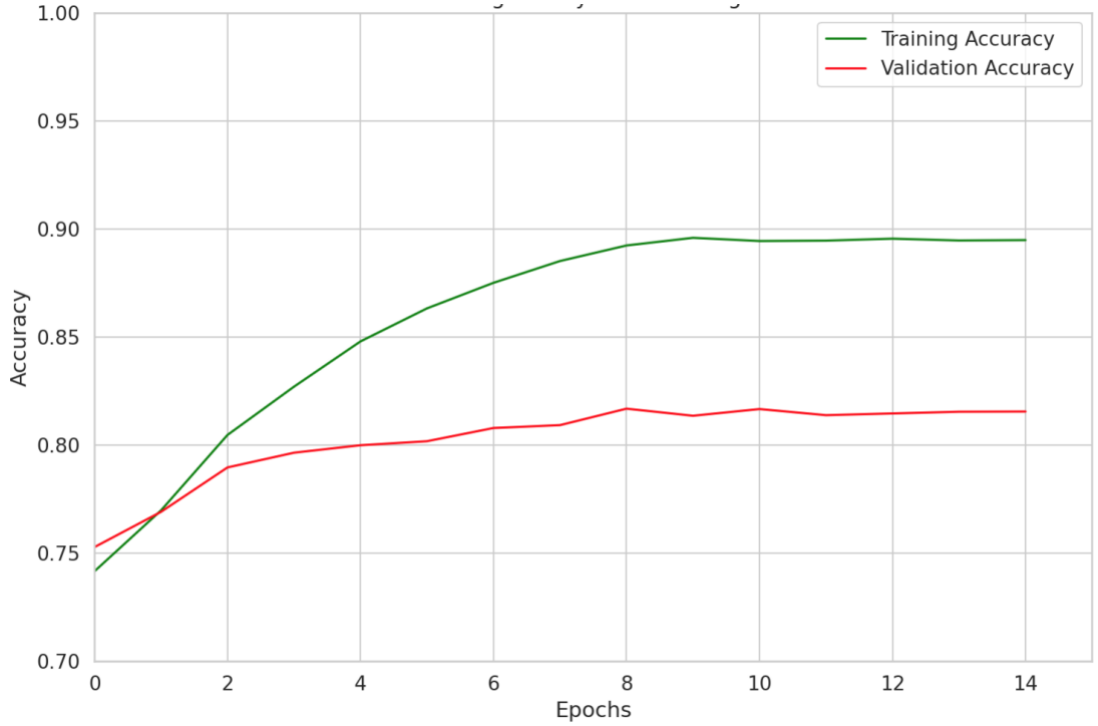

Figure 4.1-1 Training and validation accuracy of the model

Likewise, the cross-entropy loss vs. epochs graph of the model is shown in Figure 4.1-2 below. Again, looking at the graph, it is observed that the lowest loss of the lowest model is around 1.25.

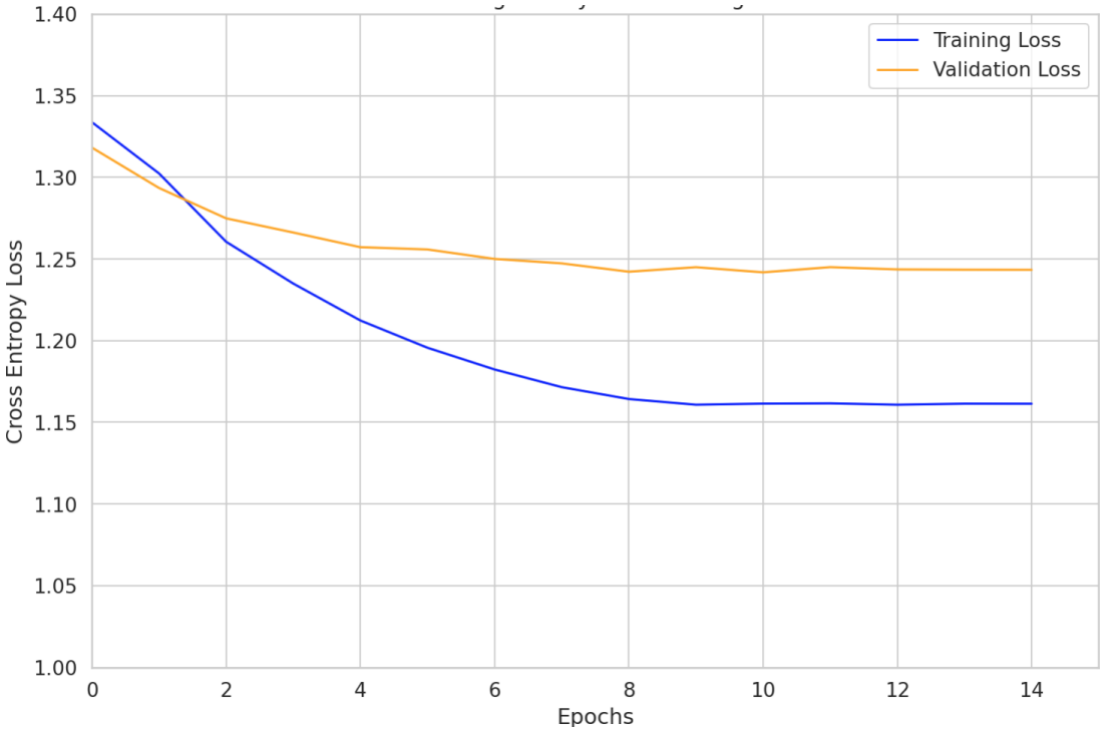

Figure 4.1-2 Cross Entropy Loss

4.2 Mobile and Web Application Results

Since the backend side is on the PC, it gives faster results than mobile, provided that the computer's processing speed depends. The technical specifications of the computer users are given below.

- Processor: Intel i3-6100U @2.30GHz
- RAM: 4,00 GB
- System type: x64-based
- Operating system: Windows

The Front-end side is cross-platform to access the application with iOS, Android, and Web (any browser). For example, the figures below show the results of true and false content detection on iOS and web applications.

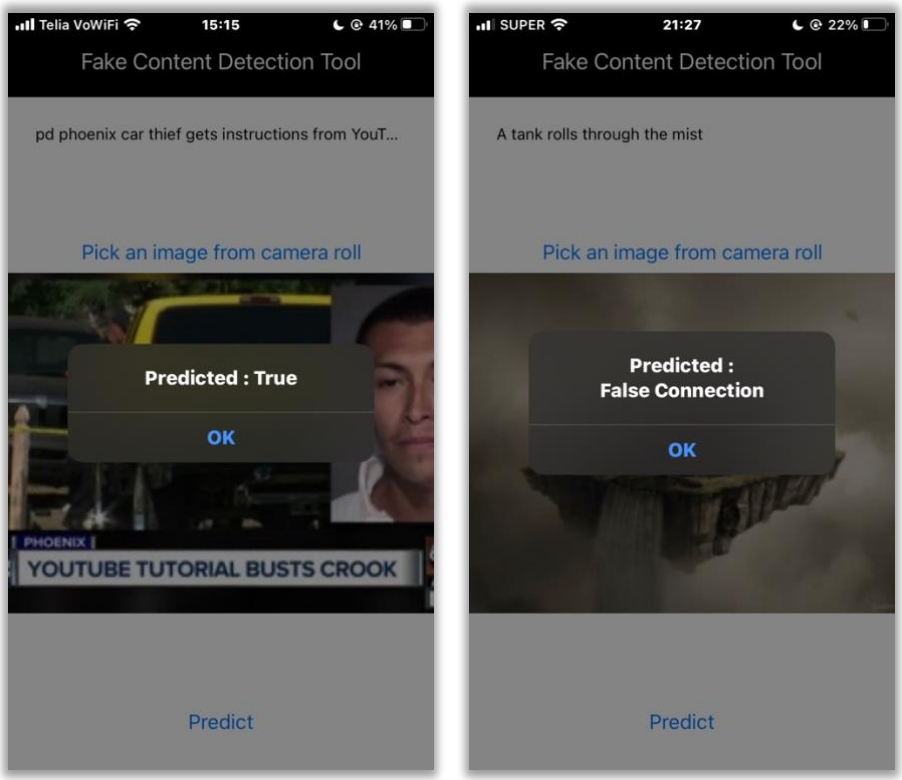

Figure 4.2-1 True and Fake content detection on iOS app

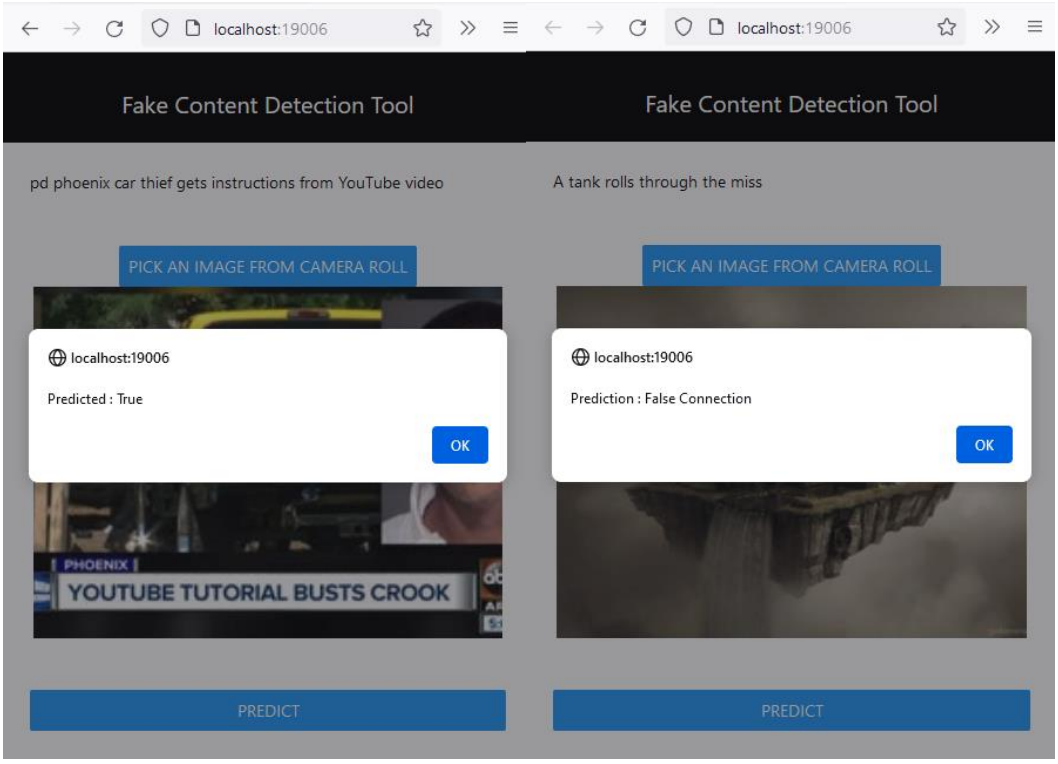

Figure 4.2-2 True and Fake content detection on a Web app

5 SUMMARY

5.1 Conclusion

This study started by conducting comprehensive research on fake content detection in social media posts. This is a crucial part for us to understand the subject. First of all, by looking at the previous studies on this subject, it was found under which sub-titles the subject was examined. The sub-titles found are forensic approaches, single-modal approaches, and multi-modal approaches. Furthermore, a summary of the most recent datasets used in these approaches is also provided. This gives researchers a great deal of information on which dataset can be used for which approach.

After the comprehensive literature search was completed, a multi-modal train with r/Fakeeddit, the most comprehensive dataset on this subject, was selected. It has been observed that this model, which was trained on the r/Fakeeddit dataset using only images and text content, has approximately 82 percent accuracy.

Later, a mobile and web application was developed using this pre-trained multi-modal. The application uses Python's Flask web framework as the backend and uses the React-Native framework for the frontend. Thanks to this application, mobile and web users can question the truth of the social media posts they suspect.

5.2 Future Work and Recommendations

The backend developed in this study can also help other applications. For this, the backend may need to be redesigned and developed as an API. After these improvements, software developers can build browser extensions, in-built social media, or micro platform services using this API. In this way, users who use such platforms will have a chance to access more accurate and faster content. In addition, since such approaches can detect spam or fishing content, they will protect users from this kind of negativities.

In the future, researchers and developers who will work on fake content detection with neural networks may close the gap in this area by concentrating on audio and video models in addition to text and image models. Further, the training time can be shortened by reducing the number of text and image models parameters.

In addition to the text and image content received from the user, the number of likes or comments that we are familiar with from social media posts can also be taken as input. A suitable model can be trained, and this model can be used in future versions of mobile and web applications. As seen from the literature review, its contribution to accuracy will undoubtedly be positive.

References

[1] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake News Detection on Social Media," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22–36, Sep. 2017, doi: 10.1145/3137597.3137600.

[2] "Many Americans Say Made-Up News Is a Critical Problem That Needs To Be Fixed | Pew Research Center." https://www.pewresearch.org/journalism/2019/06/05/many-americans-say-made-up-news-is-a-critical-problem-that-needs-to-be-fixed/ (accessed Dec. 19, 2021).

[3] B. Collins, D. T. Hoang, N. T. Nguyen, and D. Hwang, "Trends in combating fake news on social media – a survey," *undefined*, vol. 5, no. 2, pp. 247–266, 2021, doi: 10.1080/24751839.2020.1847379.

[4] B. Singh and D. K. Sharma, "Predicting image credibility in fake news over social media using multi-modal approach," *Neural Computing and Applications*, 2021, doi: 10.1007/s00521-021-06086-4.

[5] C. Pasquini, I. Amerini, and G. Boato, "Media forensics on social media platforms: a survey," *Eurasip Journal on Information Security*, vol. 2021, no. 1, pp. 1–19, Dec. 2021, doi: 10.1186/S13635-021-00117-2/TABLES/7.

[6] S. Kumar, J. v Desai, and S. Mukherjee, "Copy Move Forgery Detection in Contrast Variant Environment using Binary DCT Vectors," *Image, Graphics and Signal Processing*, vol. 6, pp. 38–44, 2015, doi: 10.5815/ijigsp.2015.06.05.

[7] J. C. Lee, C. P. Chang, and W. K. Chen, "Detection of copy–move image forgery using histogram of orientated gradients," *Information Sciences*, vol. 321, pp. 250–262, Nov. 2015, doi: 10.1016/J.INS.2015.03.009.

[8] M. Huh, A. Liu, A. Owens, and A. A. Efros, "Fighting Fake News: Image Splice Detection via Learned Self-Consistency," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11215 LNCS, pp. 106–124, May 2018, doi: 10.1007/978-3-030-01252-6_7.

[9] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, p. e1253, Jul. 2018, doi: 10.1002/WIDM.1253.

[10] R. K. Kaliyar, A. Goswami, and P. Narang, "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach," *Multimedia Tools and Applications*, vol. 80, no. 8, pp. 11765–11788, Mar. 2021, doi: 10.1007/S11042-020-10183-2/TABLES/22.

[11] Z. Jin, J. Cao, Y. Zhang, J. Zhou, and Q. Tian, "Novel Visual and Statistical Image Features for Microblogs News Verification," *IEEE Transactions on Multimedia*, vol. 19, no. 3, pp. 598–608, Mar. 2017, doi: 10.1109/TMM.2016.2617078.

[12] B. Singh and D. K. Sharma, "Image forgery over social media platforms - A deep learning approach for its detection and localization," *Proceedings of the 2021 8th International Conference on Computing for Sustainable Global Development, INDIACom 2021*, pp. 705–709, Mar. 2021, doi: 10.1109/INDIACOM51348.2021.00125.

[13] G. Bhatt, A. Sharma, S. Sharma, A. Nagpal, B. Raman, and A. Mittal, “On the Benefit of Combining Neural, Statistical and External Features for Fake News Identification,” Dec. 2017, Accessed: Dec. 27, 2021. [Online]. Available: https://arxiv.org/abs/1712.03935v1

[14] S. Singhania, N. Fernandez, and S. Rao, “3HAN: A Deep Neural Network for Fake News Detection,” *undefined*, vol. 10635 LNCS, pp. 572–581, 2017, doi: 10.1007/978-3-319-70096-0_59.

[15] Y. Fang, J. Gao, C. Huang, H. Peng, and R. Wu, “Self Multi-Head Attention-based Convolutional Neural Networks for fake news detection,” *PLOS ONE*, vol. 14, no. 9, p. e0222713, Sep. 2019, doi: 10.1371/JOURNAL.PONE.0222713.

[16] F. Qian, C. Gong, K. Sharma, and Y. Liu, “Neural user response generator: Fake news detection with collective user intelligence,” *IJCAI International Joint Conference on Artificial Intelligence*, vol. 2018-July, pp. 3834–3840, 2018, doi: 10.24963/IJCAI.2018/533.

[17] S. Girgis, E. Amer, and M. Gadallah, “Deep Learning Algorithms for Detecting Fake News in Online Text,” *undefined*, pp. 93–97, Feb. 2018, doi: 10.1109/ICCES.2018.8639198.

[18] F. Monti *et al.*, “Fake News Detection on Social Media using Geometric Deep Learning,” Feb. 2019, Accessed: Dec. 27, 2021. [Online]. Available: https://arxiv.org/abs/1902.06673v1

[19] J. Dong, W. Wang, and T. Tan, “CASIA image tampering detection evaluation database,” *2013 IEEE China Summit and International Conference on Signal and Information Processing, ChinaSIP 2013 - Proceedings*, pp. 422–426, 2013, doi: 10.1109/CHINASIP.2013.6625374.

[20] C. Boididou *et al.*, “Verifying Multimedia Use at MediaEval 2015”, Accessed: Dec. 03, 2021. [Online]. Available: https://github.com/MKLab-ITI/image-verification-corpus/

[21] Y. Wang *et al.*, “EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection,” *KDD*, vol. 18, doi: 10.1145/3219819.3219903.

[22] D. Khattar, J. S. Goud, M. Gupta, V. Varma, and J. Singh Goud, “MVAE: Multimodal Variational Autoencoder for Fake News Detection,” vol. 7, 2019, doi: 10.1145/3308558.3313552.

[23] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, “Multimodal fusion with recurrent neural networks for rumor detection on microblogs,” in *MM 2017 - Proceedings of the 2017 ACM Multimedia Conference*, Oct. 2017, pp. 795–816. doi: 10.1145/3123266.3123454.

[24] D. K. Vishwakarma, D. Varshney, and A. Yadav, “Detection and veracity analysis of fake news via scrapping and authenticating the web search,” *Cognitive Systems Research*, vol. 58, Dec. 2019, doi: 10.1016/j.cogsys.2019.07.004.

[25] S. Singhal, R. R. Shah, T. Chakraborty, P. Kumaraguru, and S. Satoh, “SpotFake: A Multi-modal Framework for Fake News Detection,” Sep. 2019. doi: 10.1109/BigMM.2019.00-44.

[26] L. Cui, S. Wang, and D. Lee, “Same: Sentiment-aware multi-modal embedding for detecting fake news,” *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2019*, pp. 41–48, Aug. 2019, doi: 10.1145/3341161.3342894.

[27] X. Zhou, J. Wu, and R. Zafarani, “$$\mathsf {SAFE}$$: Similarity-Aware Multi-modal Fake News Detection,” 2020. doi: 10.1007/978-3-030-47436-2_27.

[28] B. Singh and D. K. Sharma, “SiteForge: Detecting and localizing forged images on microblogging platforms using deep convolutional neural network,” *Computers & Industrial Engineering*, vol. 162, p. 107733, Dec. 2021, doi: 10.1016/J.CIE.2021.107733.

[29] W. Y. Wang, “‘Liar, Liar Pants on Fire’: A New Benchmark Dataset for Fake News Detection,” *ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, vol. 2, pp. 422–426, 2017, doi: 10.18653/V1/P17-2067.

[30] Gajendra Bamnote *et al.*, “AU2021106048A4 - A machine learning approach to validate the authenticity of news using natural language processing - Google Patents” Accessed: Jan. 03, 2022. [Online]. Available: https://patents.google.com/patent/AU2021106048A4/en?oq=AU2021106048A4

[31] “CN110188194A - A kind of pseudo event detection method and system based on multi-task learning model - Google Patents.” https://patents.google.com/patent/CN110188194A/en?oq=CN110188194A (accessed Jan. 03, 2022).

[32] Kai SHU, Deepak MANUDESWARAN, and Huan LIU, “WO2020061578A1 - Method and apparatus for collecting, detecting and visualizing fake news - Google Patents.” https://patents.google.com/patent/WO2020061578A1/en (accessed Jan. 03, 2022).

[33] 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”, Accessed: Dec. 01, 2021. [Online]. Available: https://github.com/bs-detector/bs-detector

[34] S. Jindal, R. Sood, R. Singh, M. Vatsa, and T. Chakraborty, “NewsBag: A Multimodal Benchmark Dataset for Fake News Detection,” 2020, Accessed: Dec. 03, 2021. [Online]. Available: https://www.theonion.com/

[35] K. Nakamura, S. Levy, and W. Y. Wang, “r/Fakeddit: A New Multimodal Benchmark Dataset for Fine-grained Fake News Detection,” pp. 11–16, 2020, Accessed: Dec. 03, 2021. [Online]. Available: https://www.journalism.org/2019/06/05/many-americans-

[36] Enzo Muschik, “Explanatory detection of fake News with deep learning,” 2021.

[37] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, “DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter,” Oct. 2019, Accessed: Jan. 03, 2022. [Online]. Available: https://arxiv.org/abs/1910.01108v4

[38] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, vol. 1, pp. 4171–4186, Oct. 2018, Accessed: Jan. 03, 2022. [Online]. Available: https://arxiv.org/abs/1810.04805v2

[39] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge,” *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, Sep. 2014, doi: 10.1007/s11263-015-0816-y.

[40] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 770–778, Dec. 2015, doi: 10.1109/CVPR.2016.90.

[41] “Welcome to Flask — Flask Documentation (2.0.x).” https://flask.palletsprojects.com/en/2.0.x/ (accessed Jan. 03, 2022).

[42] “React Native · Learn once, write anywhere.” https://reactnative.dev/ (accessed Jan. 03, 2022).

Appendix 1 – Non-exclusive licence for reproduction and publication of a graduation thesis[1]

I Şükrü BIÇAKCI

1. Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis “Fake Content Detection on Social Media Posts,” supervised by Sadok Ben Yahia and Imen Ben Sassi.
   1.1. To be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until the expiry of the term of copyright;
   1.2. to be published via the web of Tallinn University of Technology, incl. To be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.
2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive licence.
3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

03.01.2022

1 The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period.

Appendix 2 – Online Repository of the code

All of the codes of our developed mobile and web application are shared in the GitHub repository: https://github.com/subicakci/