Deep Neural Networks Towards Multimodal Information Credibility Assessment

A RESEARCH TRACK REPORT
(ISY 651/652)

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
IN
INFORMATION SYSTEMS

Submitted by:

CHAHAT RAJ
2K19/ISY/06

Under the supervision of

Ms. PRIYANKA MEEL
ASSISTANT PROFESSOR
DEPARTMENT OF INFORMATION TECHNOLOGY

DEPARTMENT OF INFORMATION TECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi college of Engineering)
Bawana Road, Delhi-110042

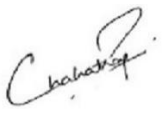
JULY, 2021

DEPARTMENT OF INFORMATION TECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi college of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, Chahat Raj, Roll No. 2K19/ISY/06 student of M. Tech., Information Systems, hereby declare that the M.Tech. Research Track Report (ISY 651/652) titled "Deep Neural Networks Towards Multimodal Information Credibility Assessment" which is submitted by me to the Department of Information Technology, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.



Place: Delhi **Ms. Chahat Raj**

Date: July 14, 2021 **(2K19/ISY/06)**

DEPARTMENT OF INFORMATION TECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi college of Engineering)
Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the M.Tech. Research Track Report (ISY 651/652) titled "Deep Neural Networks Towards Multimodal Information Credibility Assessment" which is submitted by Chahat Raj, Roll No. 2K19/ISY/06 Information Technology, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi
Date: July 14, 2021

Ms. Priyanka Meel
SUPERVISOR
ASSISTANT PROFESSOR
DEPARTMENT OF INFORMATION TECHNOLOGY

ACKNOWLEDGEMENTS

Foremost, I am thankful to God for being able to write this in full gratitude. I am immensely grateful to the Vice-Chancellor of Delhi Technological University, Prof. Yogesh Singh and the Head of Department of Information Technology, Prof. Kapil Sharma.

This thesis would not have its essence without the combined efforts of all the people involved.

I am immensely grateful to my mentor, **Ms. Priyanka Meel**, who has been extremely supportive since the beginning of my Masters. She is the lady who built me into a researcher. From the initial days until today, she has constantly been entailing and encouraging. Without her efforts, I would not have been what I am. Whilst I knew only what research is, she taught me how it is done. We share an untold understanding which was imperative for me to complete my thesis. Her words have been extremely motivational throughout the course of time. Thanks to all of her unfailing efforts!

To my professors, **Prof. Dinesh Kumar Vishwakarma**, **Prof. Seba Susan** and **Ms. Anamika Chauhan**; whose appreciations and expectations pushed me to keep climbing heights. To their worthy lectures and their out-of-the-classroom support. Thank you for showing confidence in me.

To **Mihir**, my dearest friend. He was there when I wrote my first manuscript. He's here when I'm writing my tenth! Thanks to all the proof-readings he did. Thanks to all the times he said, "Let's do this.", whenever I denied an experiment or escaped from learning something new. He was always there when I had to discuss a new research idea, when I was dwelling on a problem or when I couldn't work at all. Thanks to him for being a motivation, to his incessant endeavors and pep-talks. To **Gaurav**, my batchmate, for being there during the initial days. Your subject knowledge helped a good deal. Thanks for not helping me at times. I learned to do things by my own. Thanks to all those times you spent fixing my codes. I owe you many chocolates. Thanks to **Harshit** and **Sumit**, my batchmates, for all the good times. In addition to all the debugging you did, you always gave a good laughter. Times at the farmhouse are forever memorable, where we used to arrive to do research and end up doing nothing.
My thesis is the result of the continuous efforts of all of these people close to my heart. Research in the times of this pandemic had rather not been successful.

“It is a good thing for a research scientist to discard a pet hypothesis every day before breakfast.”

~Konrad Lorenz

ABSTRACT

False information surrounds the internet. Social media and websites are flooded with unverified news posts. These posts are comprised of text, images, audio, and videos. There is a requirement for a system that detects fake content in multiple data modalities. We have seen a considerable amount of research on classification techniques for textual fake news detection, while frameworks dedicated to visual fake news detection are very few. We explored the state-of-the-art methods using deep networks such as CNNs and RNNs for multi-modal online information credibility analysis. They show rapid improvement in classification tasks without requiring pre-processing. This thesis consists of four major contributions towards fake news research. We propose a novel ConvNet Framework for online fake news detection. Another contribution is the proposed dataset for COVID-19 misinformation detection. We develop two novel frameworks: ARCNN Framework and BERT-Multiscale CNN Framework. To aid the ongoing research over fake news detection using CNN and RNN models, we build textual and visual modules to analyze their performances over multi-modal datasets. We exploit latent features present inside text and images using layers of convolutions. We see how well these convolutional neural networks perform classification when provided with only latent features and analyze what type of images are needed to be fed to perform efficient fake news detection. We propose a multi-modal architecture that fuses both the data modules and efficiently classifies online news depending on its textual and visual content. We thence offer a comparative analysis of the results of all the models utilized over six datasets. The proposed architecture outperforms various state-of-the-art methods for fake news detection with considerably high accuracies. For the second framework, we design a novel multiscale architecture that uses two CNNs and combine them using early fusion. For multimodal detection, we combine it with a textual based feature extractor, namely BERT. The extracted features from both the models are combined using concatenation. The results obtained outperform state-of-the-art detection methods and contribute meaningfully to this research domain. Next, we formulate several hypotheses concerning fake online news and derive meaningful conclusions.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Social Media has become an important platform for the dissemination and consumption of information. Diffusion and intake of news is being widely done over Online Social Networks (OSNs). The credibility of every piece of news is not well defined and their sources are often unreliable. Users active on various social media can readily read, publish and share textual and visual information. The ease and readiness of communication contributes to increase in the spread of fake news. Majority of users share information, text and pictures without actually being confirmed if the news is true. Catchy headlines, interesting texts, pictures, videos of users' interests urge users to share such information with their peers and common interest groups.

Fake news is a piece of information that deliberately contains false information to mislead a group of people. Such kind of information is spread through traditional news media or OSNs. Such kind of false information is created and propagated to manipulate people's ideas and views, to spread political agenda or to modify business transitions. Anything that is untrue and has been proved as false is classified under fake news. Some major examples of fake news are US Presidential Elections and Pizzagate.

Types of Fake News

Here we classify fake news into 9 categories:

1. **Satire/Parody**: Humor posts, sarcasm or imitations of public figures.
2. **Mis-information**: Deliberate false information, rumors, hoaxes.
3. **Disassociation**: False connection of post with headlines, false context.
4. **Imposter**: Impersonation of official accounts and websites (e.g.: aajtak.com impersonated as aajtak.com.co).
5. **Manipulated Content**: Half-truth, forged images with manipulated headlines.
6. **Fabricated**: Hypothetical, self-created fake news.
7. **Advertisements/PR**: Ads with fake claims to attract people.
8. **Clickbait**: Links redirecting to false information pages.
9. **Propaganda**: Misleading news promoting political causes.

Figure 1: Types of Fake News

Fake News Motives

The reasons why people spread fake news range from the idea to harm someone to simply fun. The motive of transmitting fake news depends upon the actor i.e. the person or entity that is spreading it. Here, we present five broadly classified intents or motives of why people spread fake news:

1. **Malicious Intent**: To defame a person, entity or an organization.
2. **Political Influence**: To manipulate people's ideas towards or against specific political persons and parties to alter the outcomes of elections.
3. **Profit**: To increase monetary gains and enhance business.
4. **Popularity**: To increase traffic on specific websites and accounts, become famous.
5. **Fun**: To create humor and entertainment.

Fake News Actors

Users of media platforms who unconsciously or maliciously (with intent) spread fake news are termed as fake news actors. We provide types and subtypes of such actors who in any manner contribute into spread and increase in fake news on various platforms.

Figure 2: Fake News Actors

Fake News Impacts

Users of media platforms who unconsciously or maliciously (with intent) spread fake news are termed as fake news actors. We provide types and subtypes of such actors who in any manner contribute into spread and increase in fake news on various platforms.

1. **Political Effect**: Changes outcomes of elections, hampers democracy, disrupt normal political proceedings.
2. **Fear**: Fake stories about situations incorporate fear in minds of citizens like fake news about occurrences of natural disasters- tsunami, earthquakes etc. might create havoc in general public.
3. **Racist Ideas**: Fake news about societies that promote communalism and racism and lead to social discrimination towards a specific societal group.
4. **Image Violation**: Disregards status of celebrities, political leaders, business organizations and brands.
5. **Riots**: Creates war-like situations between states, nations, communities. For instance, Hindu-Muslim uprisings.
6. **Stock Variations**: Fake news about particular brands alter the sale and purchase of items and hence change their stock prices. Fake news causes a great impact in stock market.

7. **Mislead**: Fake news about vacancies, advertisements of products mislead people into joining fraudulent companies, buying wrong products affecting people's lives in several ways.
8. **Health Issues**: Fake medicinal vitae posted on social media by people not certified in medicine cause health issues to people who follow and act on such content.

Figure 3: Fake News Impacts

The world of information is shrinking into an online space. From amongst 4.4 Billion internet users, 68% people get their news from social media rather from traditional news media. This news could be shared by authentic news media or by general public. News on OSNs appears to be relatable and interesting to the age groups (12-30) of most users. Consuming news from social platforms is much easier and feasible. It is easily and directly approachable. It equips people with diversity of news by just a tap of a finger, without dedicating extra time to approach to traditional media news. Social media provides a platform to debate. While in traditional news mediums, people can just read, view or listen to news, OSNs provide users also to react and express their opinions and emotions towards the news. It is in short an interactive platform. It stays as the restraint of users to verify using their conscience, the legitimacy of the post. Users need to keep an open mind to perform such classification. Cognitive Bias, Continued Effect and Selective Exposure are the limitations in classifying information as true or fake.

Social Media such as Facebook and Instagram provide users with posts according to their interests. People are exposed to news and posts only from pages they follow and the accounts they have befriended. This limits the users' exposure to information that coincides with their faiths and beliefs. Users avoid any posts that contradict their preset views. To confirm the legitimacy of a post, users usually rely upon the number of followers of the account sharing the information, number of likes, shares and comments on the post. The nature of comments and reactions to a post allow discerning its authenticity.

Users share what they believe to be true, not what is actually true. Fact-checking before sharing is necessary. A thing of concern is whether bots or malicious users are posing as authentic news sources to spread propaganda or misinformation. Users do not want to input their efforts to grab news but would consume it if it is readily available in front of them without investing much of their time. They would surely read a catchy headline that pops up on their mobile phones but would not care to go through the entire article. This prevents them to know the factual information beyond the headlines and they do not fact-check the authenticity.

Visual data attracts viewers more quickly than words do. A Human brain captures and rapidly analyzes a news item and often flags it as fake or real by just a glance of its title, image, or a small segment of it, mostly without going through the entire textual content. It does this based on the preexisting knowledge in our conscience. Even if it does go through a whole text, there are very few references and not enough time to check for the authenticity of the content we come across. Various content creators exploit these drawbacks of the human brain and behavior. There is a need for technological state of the art methods to assess the credibility of content, textual or visual, and authenticate it as fake or real. Online media emerged as a platform to share ideas, views, news. With the advancement of mobile devices and the internet, news became easily accessible to people who were either deprived of or uninterested in official news sources such as television and newspapers. The long and seemingly tedious to read texts became easy to understand as images and videos now accompany them. In the same process, it also became challenging to detect the truth in such content.

In the present scenario, online media is losing its charm and credibility as content creators lure users to gain popularity and money using the content they post online. In this process, they do not pay heed to the authenticity of the information, ignore the verification process, and mix up misleading or tampered images or clips with the texts. Content creators focus on posting catchy and attractive content that bags them many likes, comments, and

dollars. Sometimes both the text and graphic content are intentionally made erroneous to spread fake news, making the entire content even more unrealistic. Hence there is an urgent need to design and develop a new classification method to assess the credibility of content, textual or visual, and segregate it as fake or real. If textual and visual factors are taken collectively, fake news detection methods have proved to provide higher accuracies than unimodal detection methods. Machine learning and deep learning-based detection mechanisms depend on fake or real news by analyzing the text's features and visual data. Users consuming information play an essential part in stopping the spread of fake content at the root level or circulating to reach a great mass affecting political, social, and economic lives. The algorithms so far used depend upon news data collected from websites and social media platforms, which are later classified into binary (real and fake) or multiple (ranging according to their severities) labels by crowdsourcing or third-party authenticators.

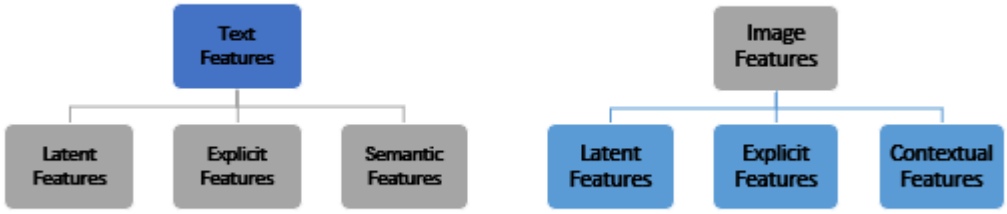


Figure 4: Types of Textual and Visual Features

With the advent of massive data and news content online, the intricacies add up when multiple data forms are available. Despite being beneficial in terms of easy transmission and news consumption, multi-modal data also presents a strenuous task for detecting fake news amongst them. The modalities prevalent on online media include text, image, audio, video, and hyperlinks. With the vast accompaniment of text with visual data, the effectiveness of news rises. A large amount of visual data makes verification difficult as multi-modal data does not guarantee the credibility and attracts more attention than pure text contents. Multi-modal features are expected to be more beneficial in detecting fake news as compared to unimodal features. Few of the excellent quality datasets available for scientific research include binary labeled datasets and multi-label datasets such as Mediaeval, Sina Weibo, PolitiFact, Emergent, and Resized_V2.

We propose that online social media images consist of three features: latent features, explicit features, and contextual features. Latent features are extracted using layers of

convolutions. Deep convolutional networks are capable of learning kernel values that are utilized to extract latent features. According to Yang et al., explicit features are hand-crafted features such as the resolution of an image and the number of faces in the picture. Apart from these two intrinsic features, contextual features are based on semantic relationships between the text and the image. We have executed convolutional neural networks for text and image classifications. CNNs provide an advantage to extract features directly from raw input without any pre-processing required. CNNs reduce input data on various layers such that only required information is preserved and worked upon to make essential predictions.

As we have a tendency to observe trends among faux news detection, an enormous quantity of printed work and resources are out there for text-based detection. tidy amounts of frameworks and detection mechanisms victimisation matter options are designed since the matter domain emerged. Machine learning and deep learning algorithms have been applied mostly to supply solutions thanks to their extreme quality during this domain. These embrace sentiment analysis, text mining, stance classification, similarity analysis, and so forth Texts are analyzed supported their sentence structures, words, punctuations, tone, grammar, and pragmatics. matter fake news analysis is that the domain well explored and worked upon by an outsized number of researchers. numerous different detection techniques have emerged, like victimisation cognitive psychology features, deception modeling, faux news spreading prediction, graph-based propagation detection, and name scores.

Fake news spreads within the kind of matter or visual data. we tend to aim to elaborate on however mechanisms are developed to traumatize every type of pretend data that spreads textually through pictures or videos. Towards pretend visual news, tasks performed are relatively lower in number. The domain started gaining attention in 2013 and rose significantly from 2017 to 2020. Image fake news detection started gaining importance once image-accompanied news and posts started showing at giant on on-line platforms. the increase in visual knowledge created ways that for fake news to feed in and so inspired researchers to explore this domain. Cao et al. have highlighted the role of exploring visual data whereas police work fake news. pretend news detection has shown notable performance enhancements once a visible analysis is combined with text. Most of the visual classification techniques applied neural networks that provided fast and economical results. pretend pictures is classified as tampered images wherever a number of the opposite manipulations are created within the image or as dishonest images where the context of stories content and image don't comply. Some fake

news is additionally in the course of older images from other events, i.e., combining recent news with out-of-date images. Another class of faux images that recently emerged is computer-generated images normally created by Generative Adversarial Networks (GANs). of these sorts of pictures contribute to pretend news. Any FND framework has not been ready to discover fake news that revere to any or all these varieties collectively. Individual modules are created for various tasks. Some architectures will classify tampered images well, whereas others can spot dishonest images that don't match the context. Some frameworks classify images supported applied mathematics options of matter and visual data. Parikh and Atrey provided a survey for transmission FND. there's an absence of a wholesome framework that might expeditiously detect fake news below all modalities.

CHAPTER 2

LITERATURE REVIEW

Since the advent of online social platforms, the need to establish baselines for information-assessment across all information flow channels has demanded researchers' attention. It has been dubious how wisely people using these platforms actively communicate and digest the information circulating on the internet. With the exponential emergence of social media as a globally used platform in the recent decade, we have encountered massive escalation in fake news dispersion. Any act of deliberate, miscreant, or unverified inclusion of information creates fake news. Substantial instances include widespread misinformation since January 2020 claiming World War 3 [1], its probability, countries involved, and tentative dates. The arrival of the COVID-19 pandemic has led to a multi-fold rise in disseminating fake news globally. This escalated spread of misinformation amidst the pandemic has been termed as 'Infodemic.' Statements such as "Turmeric powder and black pepper cure coronavirus [2]", "Cocaine treats COVID [3]" etc. have appeared as post contents on every social site in the forms of text messages, images, and videos. Such fake news can be found listed in various official websites debunking the claims [4]. Another fake message states a WHO-issued four-step protocol to prevent COVID-19 circulated at large on most online social networking platforms, especially on WhatsApp [5]. Conspiracy theories and tweets conveyed that the stated countries created the novel coronavirus as a weapon to ignite bio-war [6]. Subsequently, fake news about the pandemic became as contagious as the virus itself. A huge source of fake news was the 2016 US presidential elections that have also focused on many fake news researchers. Other such events and examples of fake news are Pizzagate [7], Indian and Brazilian elections [8], Hurricane Sandy [9], spying technology in Rs. 2000 Indian currency notes [10], Citizenship Amendment Act 2019 [11] & Article 370 (Kashmir, India) [12], etc., which have had huge adversarial impacts on the lives of people.

Figure 5: Visual fake news example of a shark in Houston (Source: USAToday)

Social media interactions have contributed a lot towards inventories of big data. Users accord it in text, image, video, audio, emoticons, reactions, etc. Text, images, and videos compose a major actuating portion of the information on the internet and hence blot the grey areas of fake news. Since fake news and its critical nature came into the picture, several scientific developments have been made in textual fake news detection. Fake news detection technologies using NLP, text classification, vector-space models, rhetoric structure theory, opinion mining, sentiment analysis, graph theory, deep neural networks, and others have been created, reviewed, and summarized by fellow researchers [13-16]. Probing of image fake news is comparatively lower, and of videos, it is negligible. Most of the information users encounter on the internet is accompanied by visual representations, either using an image, video, or other modalities. Visual data is quickly gazed upon and leaves a lasting impact. It is the breeding ground of tampering, manipulations, and forgeries. With technology emanating, editing applications and techniques, images and videos are being meddled to mislead information consumers. Moreover, online social media allows users to add to their data, be it real or fake. Recent examples include an image of a shark on a Houston freeway that went viral during Hurricane Harvey. The image was shared and retweeted at large, creating havoc amongst people. Misinformation has been so outrageous that even US President Donald Trump could not escape from it and went on to retweet a fake video that stated anti-malaria drugs could cure coronavirus, which was later brought down when proven fake [17]. In 2020, during the coronavirus pandemic, fake news claiming Kim Jong Un dead, absconding, or assassinated spread widely [18]. Fake videos of his funeral were shared on social media.

Figure 6: Outdated image of King Jong Un's Death (Source: nypost.com)

We broadly categorize data modalities into text, images, and video that are spreading fake news on the internet. As we observe trends among fake news detection, a huge amount of published work and resources are available for text-based detection. Considerable amounts of frameworks and detection mechanisms using textual features have been designed since the problem domain emerged. Machine learning and deep learning algorithms have been applied largely to provide solutions due to their extreme popularity in this domain. These include sentiment analysis, text mining, stance classification, similarity analysis, etc. Texts are analyzed based on their sentence structures, words, punctuations, tone, grammar, and pragmatics. Textual fake news analysis is the domain well explored and worked upon by a large number of researchers. Various other detection techniques have emerged, like using psycholinguistic features, deception modeling, fake news spreading prediction, graph-based propagation detection, and reputation scores.

Fake news spreads in the form of textual or visual data. The fake news domain has been expressively presented by Sharma and Sharma [19], Figueira and Oliveira [20], Torabi and Taboada [21], Zhang and Ghorbani [22], Tandoc et al. [23], and Shu et al. [24]. Summaries of noteworthy works have been provided in reviews by Mosinzova et al. [25] and Rubin et al. [26]. In contrast, Rajendran et al. [27] brought into the light the importance of Deep Neural Networks for stance classification.

Moving towards fake visual news, tasks performed have been comparatively lower in number. The domain started gaining attention in 2013 and rose considerably from 2017 to 2020. Image fake news detection started gaining importance when image-accompanied news and

posts started appearing at large on online platforms. The rise in visual data made ways for fake news to seep in and thus encouraged researchers to explore this domain. Cao et al. [28] have highlighted the role of exploring visual data while detecting fake news. Fake news detection has shown notable performance improvements when a visual analysis is combined with text. Most of the visual classification techniques applied neural networks that provided quick and efficient results. Fake images can be classified as tampered images where some of the other manipulations are made in the image or as misleading images where the context of news content and image do not comply. Some fake news is also accompanied by older images from other events, i.e., combining recent news with outdated images. Another category of fake images that recently emerged is computer-generated images commonly created by Generative Adversarial Networks (GANs). All these types of images contribute to fake news. Any FND framework has not been able to detect fake news that revers to all these types collectively. Individual modules have been created for different tasks. Some architectures can classify tampered images well, while others can spot misleading images that do not match the context. Some frameworks classify images based on statistical features of textual and visual data. Parikh and Atrey [29] provided a survey for multimedia FND. There is a lack of a wholesome framework that could efficiently detect fake news under all modalities. As we talk of visual data, fake news videos have taken up at large over social media. These have a great impact on the minds of viewers and bring adversarial social and political effects. Frameworks for video fake news detection are very few. Multiple features are needed to be studied for video analysis, being a complex task containing spatial and temporal features, speech, and movements. Few researchers have applied it using inconsistencies between speech and lip movements; few have attended analyzing facial expressions following similar trends in real and fake videos, while few utilize the image information at every frame. Other multimedia news in audio, podcasts, and broadcasts are yet less infiltrated with fake news. With an acutely low amount of work performed, video fake news detection is an emerging problem domain that researchers need to pay heed to.

2.1 CHALLENGES IN MULTIMODAL FAKE NEWS DETECTION

The domain of fake news detection gained popularity very quickly. Large amounts of unverified and uncredible posts have been misleading people. Using linguistic features for credibility assessment of content is a popular and widely-used method. Here, we list the existing challenges to detect fake news spreading through all types of data.

1. **Visual Fake News Detection:** The fake news menace is rising. It began with spreading through text and has now started gripping users through all forms of multimedia. Visual data probable to fake news exploitation can be categorized into images or videos. There are various existing approaches to detect text-based fake news. However, visuals play a great role in impacting viewers' minds and therefore are being infiltrated with fake news in the current generation. An image or video can be easily modified using media editing applications. Various manipulations in visual data go unrecognized through the viewers' eyes. It is very difficult for humans to observe minor changes in modified images and videos to classify them as fake or real. Automated tools are required which can identify minute variations created made to fabricate or manipulate visual data. This poses a great challenge for researchers in designing visual-based fake news detectors.

2. **Auditory Fake News Detection:** Auditory fake news is a type of fake news that has been in existence but is going unnoticed. Many social applications allow users to share recorded audios. These audios files are vulnerable to spreading fake content, propaganda, unverified information, and more. This type of multimedia has not been put into use for credibility assessment. There are no fake news detection mechanisms that incorporate audio as a sole modality or in combination with other data modalities. This issue needs to be addressed to prevent the contamination of audios with false content and serve its early detection.

3. **Detection of Embedded Fake Content:** When one type or modality of data is fixed or embedded within another type of data, it is embedded content. A new type of social media posts is spreading widely known as a 'meme.' It is an image or a video, mostly with text embedded on it. Various forms of media like text, image video, gifs, or hyperlinks are embedded into other forms. It is a complex task to analyze media that is embedded in some other data type. There is an upsurge of fake content in the form of embedded media or memes. Efficient detection mechanisms are required to fight such misleading content.

4. **Multimodal Datasets:** To build fake news detection mechanisms, most machine learning and deep learning tasks require large amounts of data. There is a lack of real-world multimodal datasets. Text-based fake news datasets are more in number than visual or multimodal datasets. Lack of proper datasets limits the extent of research. There is a need to collect real-world fake news data that consists of various types of information like text, image, video, and meta-data.

5. **Holistic Detection Mechanism:** The research community has encountered many techniques that can robustly detect fake news. These techniques use linguistic features, visual features, sentiment scores, social context, network/propagation-based features, meta-data, and hybrid features. There is no such mechanism at present that extracts all these details from a given fake content and predicts its integrity based on all the contributing factors. Different researchers have highlighted the importance of all of these techniques individually or in hybrid combinations. It is worthwhile to consider all these features for building a holistic fake news detector.

6. **Real-Time Verification:** Provided information-spread ease through the internet and online social platforms, fake news is being generated and spread at every instant. Fake news can be about anything and anyone. It spreads continuously as we interact on the internet. Existing detection tools either require users to self-validate a piece of news by fact-checking on their website/application or classify news late after it has been spread and affected various aspects of life. The world needs a system that analyses content in real-time and declares it as fake or real based on its decision.

7. **Lack of Literature:** There is a lack of significant literature in the domain of multi-modal fake news detection. Although many authors have presented textual detection mechanisms, work done in the multimodal domain is minimal and not complete. In this work, we endeavour to cover all the past research performed using multiple data modalities other than only text-based techniques. We highlight works that have used visual content, alone or along with textual content, for detecting fake news.

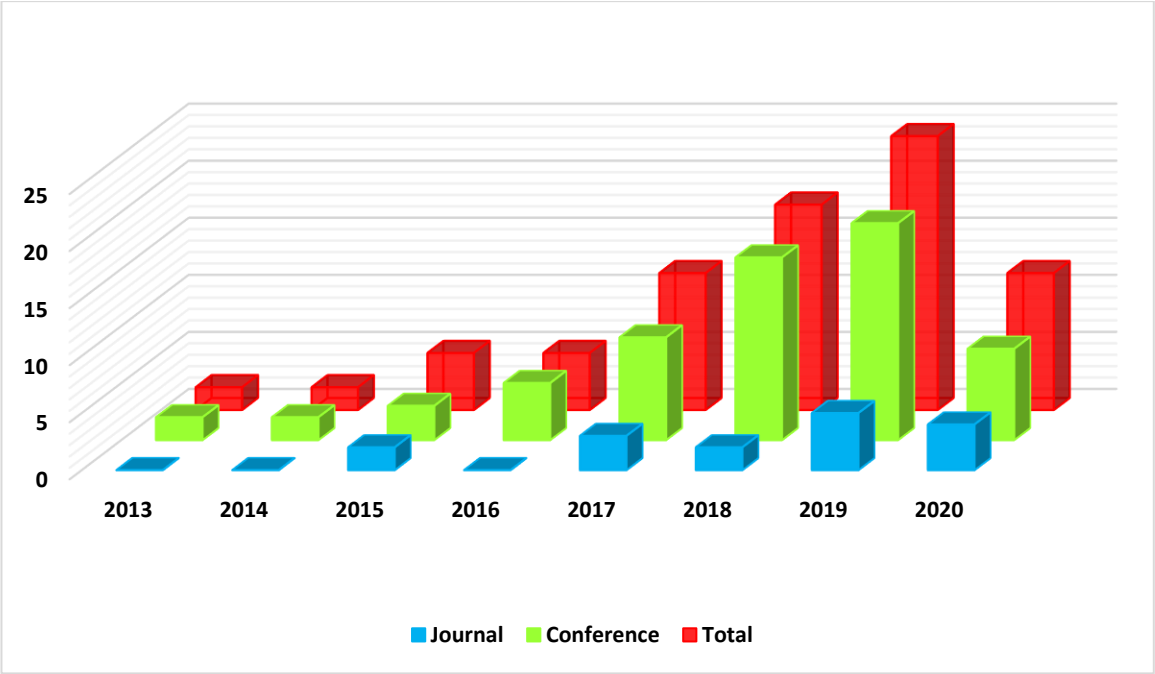


Figure 7: Year-wise trend of published work

2.2 FAKE NEWS MODALITIES

Fake news is defined as any piece of false information that misleads people. It can be deliberate, fabricated, or simply unintentional. The intent of spreading false news could be maliciously intentional, political, for gaining monetary benefits, popularity, or simply for fun. While referring to data modalities, fake news spreads through text, images, videos, audios, hyperlinks, embedded content, and hybrids. Because of less or no work in the remaining modalities, research is limited to textual and visual modalities.

Text: This is the most popular mode of communication on the internet. People interact through textual matter on social media platforms, websites, blogs, e-mails, personal messaging, and more. Most of the false information spreads through text on the internet. Fake news is found propagating on social media posts, articles, and online messaging services. Text is the simplest and most used way for an internet user to convey his concerns. Being a largely used modality for communication, it also accounts for a large amount of fake news. The screengrab is taken from Twitter. In the tweet, the user falsely attributes a claim to WION News, which states that China is hiding the real numbers of death amidst the coronavirus pandemic. The post says that SO2 concentration around Wuhan, China, has grown due to the burning of a large number of dead bodies. The claim is false and has been debunked by various fact-checking websites.

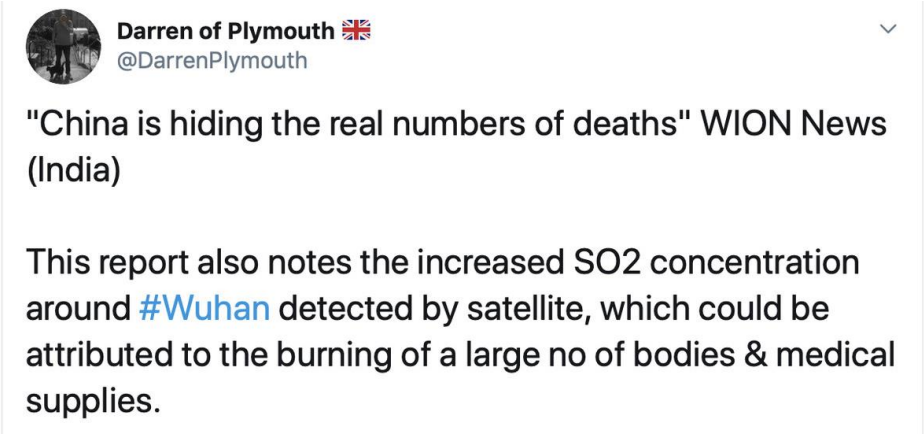


Figure 8: Textual fake news example (Source: Twitter)

Image: An image is a visual representation of something. Images are highly vivacious and impactful for depicting anything. They leave lasting impressions on the minds of viewers, whereas words can be forgotten shortly. They have suddenly gained popularity with the increase in the feasibility of sharing them. Images go through certain manipulations to carry a false message. The use of photo editing tools supports these manipulations. Some examples of editing techniques are cropping, splicing, copy-move, retouching, or blurring. Any image can be manipulated to convey a false message, which contributes to fake news. Often, they are not manipulated but accompanied by false text. Many of the times, irrelevant or out-of-context image is placed with fake text. All of these types of images emanate false conceptions accord with fake news. A Facebook post shows a girl rescuing a koala bear from Australia's bushfires. Originally, the picture is a digitally created artwork used out-of-context to match the bushfire situation.

Figure 9: Example of fake news image (Source: Facebook)

Video: Sharing of videos online became immensely easy with the introduction of YouTube. The platform allows the feasible sharing of information through video content. Social media platforms allow various techniques for communication through videos. This can be done in the form of regular posts, stories, ads, or even comments. This viability of video interaction gives room for sharing of fake content through videos. Videos are a powerful and impactful tool. They are capable of successfully manipulating people through their content. Therefore, it raises a serious concern to authenticate video content and decide whether a video is credible or not. Figure 10 shows a screengrab from Twitter that shows a video claiming Vladimir Putin's daughter was getting the first shot of coronavirus vaccine. The girl is, in fact, a volunteer and not the Russian President's daughter.

Figure 10: Example of fake news video (Source: Twitter)

Other Modalities: Numerous data types have not been analyzed for fake news yet due to limitations of exposure and datasets. These include embedded data types, audios, and hyperlinks (clickbait). Embedded media is that where one type of data is merged or superimposed onto another. For example, textual matter on images or videos, embedding audio in images, altering audio in videos, etc. Detection of fraudulent content in such a data type is complex and challenging. There is a lack of past research in this area. Figure 11 shows a meme with a text embedded on it that says that the North Korean leader Kim Jong Un faked his death to expose traitors. Many such false statements and claims circled the internet. Various fact-checking sites have debunked these claims.

Figure 11: Fake example of a text-embedded image (Source: Facebook)

2.3 DETECTION MECHANISMS

Various fake news detection algorithms utilized by researchers are explained below. Figure 12 depicts the percentage distribution of used algorithms and methods in the reviewed articles. These include Deep Neural Networks, Image/Video Forensics, explicit features, and other methods. Textual detection mechanisms have employed several machine learning and deep learning classifiers. They have been successful enough in this domain. For visual analysis, the most popular tools are Deep Neural Networks. Many researchers have combined these with the use of explicit features available on the web. These can be statistical features, user-based features, post-based features, propagation features, or more. Another popular method is the use of reverse image search on search engines. This method is useful in identifying the integrity of a particular image. Tools are available that use this mechanism to allow the verification of fake news. Several other methods of image and video manipulation have started being merged with fake news detection. This has bridged the gap between methodologies and brought the problem domain and possible solutions in a common range. The applications, advantages, and disadvantages of the following methods have been summarized in table 1.

Convolutional Neural Network

First introduced in the 1980s, CNNs have come a long way in the domain of computer vision. They have been applied to Natural Language Processing [30], image classification [31], video classification [32], object recognition [33], time-series forecasting [34], anomaly detection [35], speech analysis [36], handwriting recognition [37] and the likes. For image classification, CNNs require training over large image datasets. Their learning process occurred to be substantially faster than previous methods known, an underlying feature that brought

CNNs into the picture. They are efficient in analyzing latent features present inside an image or a video. A rich survey of the latest Convolutional Neural Networks has been provided in Khan et al. [38]. As compared to images, little work has been done in video classification using CNNs as videos are more complex to process owing to their temporal dimension. Most works utilize CNNs to classify videos by extracting images at every video frame [39]. Another method treats spatial and temporal domain separately and classify them using two convolutional neural networks and fusing them after that. Many researchers have also applied CNNs for text classification using one-dimensional convolutional networks.

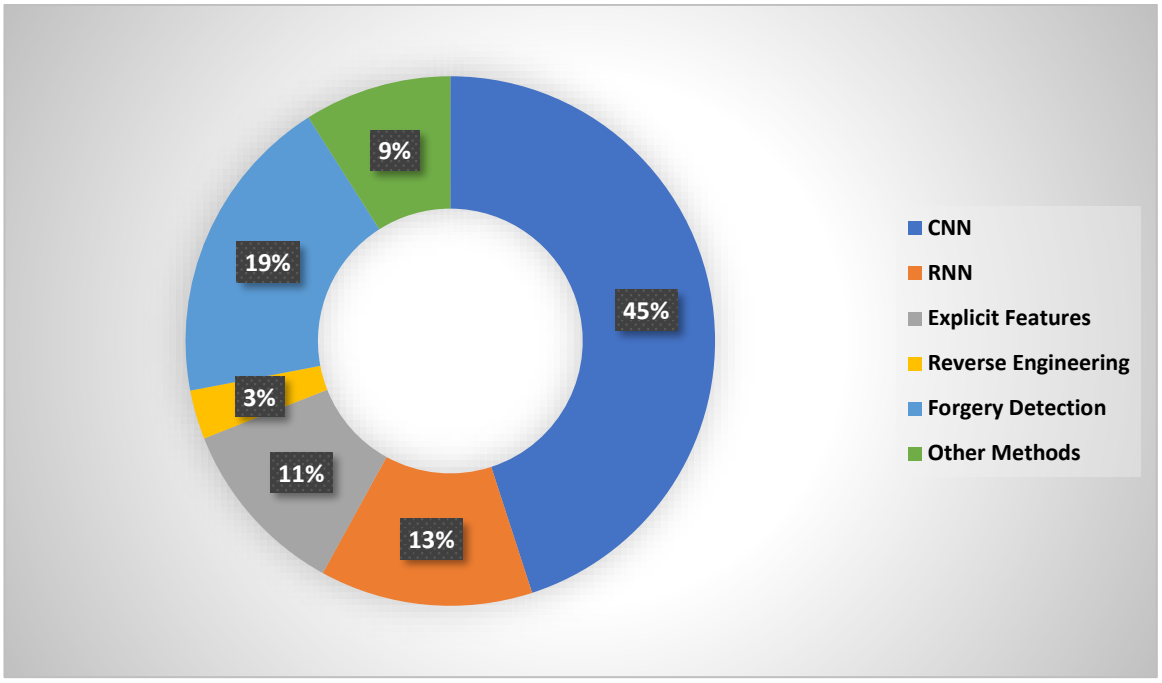


Figure 12: Distribution of reviewed articles representing each method

Recurrent Neural Network

Recurrent Neural Networks analyze sequential inputs like text, image, speech, video, and output in a feedback loop. A network with feedback loops is created, which allows RNNs to retain information and train themselves. RNNs have been utilized in image classification [40], video classification [41], object recognition [42], video annotation [43], time-series prediction [44], anomaly detection [45], sentiment analysis [46], speech recognition and other ML and DL tasks. LSTMs (Long Short-Term Memory networks), a special case of RNNs, have been widely utilized in fake news classification tasks in multiple modalities. These train much faster and can perform complex classification tasks than other RNNs. 13% of the articles reviewed in this work performing multi-modal fake news detection have utilized RNNs and a combination of RNNs and CNNs.

Table 1: Comparative analysis of fake news detection methods

Domain	Method	Applications	Advantages	Disadvantages
Deep Learning	Convolutional Neural Networks	Image, video, object, speech, time-series, handwritten data recognition, text mining [30-37]	Adaptively learn classification features	Cannot detect image manipulation on its own Require large labelled datasets
Deep Learning	Recurrent Neural Networks	Time, sequence-based predictions, Text, speech, image, video processing, object recognition [40-46]	Great memorizing capacity, usable with CNNs	Slow computation, difficult training, hard to process long sequences, exploding and vanishing gradient
Reverse Engineering	Reverse Image Search	Image origin tracking, metadata tracking	Allows verifying variations in images by matching with originally occurring images and text	Can work with images previously present, cannot detect every type of fake image
Explicit Feature Analysis	Statistical Analysis, Semantic score analysis, User-profile feature analysis, Propagation feature analysis, Geolocation, Psycholinguistics	Explicit feature detection in all modalities based on features present out of the content	Provides prominent non-data features, measures semantics and relevance	Does not include latent features

Data Forensics	Image/video tampering detection, Face manipulation detection	Forged image and video detection, face-swap detection, deepfake detection, editing detection	Easily detects changes made in an image area by editing, removal, addition etc.	Difficult to detect minute manipulations

---

Reverse Engineering

A popular method for fake news detection is to look up the content in question on the internet. Post's credibility is assured by matching it with the occurrences that appear in the search results. The methodology used under reverse engineering for fake news detection is Reverse Image Search. Search engines like Google, Yahoo, and Bing allow their users to input a query image and provide them with relevant information about the image. This process is utilized in fact-checking the authenticity of an image. We can get to know how old an image dates back and where did it appear first. Metadata can also be extracted from such visual data. It also helps us verify the context of the image to the text it accompanies. This method is used by automated fake news detection tools, applications, or web plugins.

Explicit Features

Under explicit features utilized for FND tasks, we categorize statistical features (no. of words, likes, shares, retweets, comments, reactions, etc.), similarity features that analyze the similarities between content and visual information of a news article and state how well both of these are correlated, semantic features that verify meaningfulness of data, user profile features that provide information about users' age groups, backgrounds, faiths and beliefs, inclination, online social behavior, and other relevant profile information, propagation features that help analyze the flow of fake news among networks and people, geolocation features those study areas of fake news generation and propagation and other external features. These features, when combined with other modalities, increase the weightage of detection accuracies. They serve as an important factor for fake news analysis and detection.

Data Forensics

Images and videos, given the current technological advancements, can be easily edited and tampered with. We have classified fake news detection techniques using forgery detection,

splice-detection, copy-move detection, face-swapping detection, face manipulation detection, pixel-based forgery detection, photoshop detection, object-removal detection, repurposing detection, and other similar editing detections under image forensics. This method verifies the credibility of images and videos without a requirement of their original version. Algorithms utilized can detect manipulated regions in images and videos.

Other Methods

Few other methods that have been utilized by researchers for fake news detection can be named as co-occurrence matrices, blockchain, pattern recognition, etc. It has become popular to match the semantics between post text, image, and video. Few of the latest works verify if the post's modalities convey the same meaning and then classify them as real or fake. They have provided a new dimension to investigate fake news detection. This area provides opportunities to be explored more, enhance currently available methods, and leverage new ones.

2.4 RELATED WORKS

Qureshi and Deriche [47] explained the taxonomy of types of forgeries found in images: copy-move forgery, image retouching, resampling, image splicing. They have also discussed pixel-based forgery detection methods in images that include contrast enhancement detection, sharpening filtering detection, median filtering detection, resampling detection, post-processing editing detection, copy-move detection, and image splicing detection. Brezeale and Cook [48] provided a survey of existing video classification methods that classify videos using text features, audio features, visual features, and combinations. Boididou et al. [49] have reviewed various methodologies for classifying multimedia data on Twitter that include verifying cues, assessing the source and user credibility, content credibility, image forensics, verifying multimedia use task and have described the verification approaches used, namely UoS-ITI, MCG-ICT, and CERTH-UNITN. Anoop et al. [50] deeply studied fake news detection methods on textual modality, image modality, network modality, temporal modality, and knowledge-based approaches. They have also discussed popular datasets for use. Tolosana et al. [51] reviewed existing face-manipulation detection methodologies. Saini et al. [52] neatly summarized supervised, semi-supervised, and unsupervised multi-modal FND frameworks that include baselines like MVAE and EANN. They have compared state-of-the-art by nicely tabulated data.

Image tampering has become easier than ever, given the advances in photo editing tools. It is crucial to detect such forgeries to keep a check on fake news data. Fake images accompanied by fake news are categorized as in figure 13. The categories are: tampered/edited images, outdated images used with a later situation, and images out-of-context with the accompanying text. Figures 14, 15, 16, 17 show examples of fake news images. Approaches utilized to detect these fakes include supervised deep learning techniques that require huge training data. Deep neural networks have been successful in the classification of manipulated images from the originals. Table 2 summarizes all the tasks related to fake news detection that involve visual modalities. It helps to understand the necessary details of the related works easily. In table 3, a summary of supportive works is provided, which utilize data forensics mechanisms to identify tampered visual data.

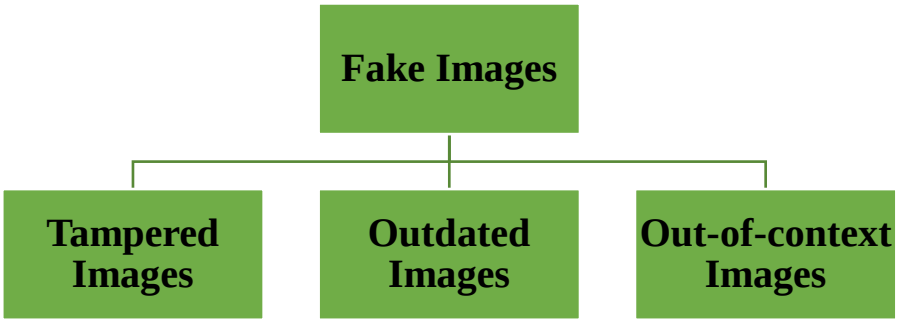


Figure 13: Types of fake images

Elkasrawi et al. [53] allowed users to verify images present online using metadata and feature analysis automatically. The approach has two phases: first involves checking the image for its authenticity, whereas, in second, it is verified if the image has any alterations. The first phase attempts to find other occurrences of an image on the web using Google Image Search. This results in the appearance of the query image in its similar or manipulated versions. Along with several versions of the image, metadata like URL of the image, article in which it appears, publishing date, time, thumbnail, etc., are retrieved. The image's authenticity is validated by matching the timestamps of the query image and search result image. If the image dates back from the date of the query image, it is regarded as fake. In several occurrences, the algorithm uses k-means clustering for the resulting images and their timestamps. In the second phase, image matches are retrieved, and alteration detection is performed. This is done using edge comparison and checking image alignment. The algorithmic features are conglomerated and deployed in a Chrome browser extension for image-based fake news verification.

Wang et al. [54] built a neural network framework for incoming real-time events. This Event Adversarial Neural Network (EANN) framework can handle event-invariant features, thus allowing the detection of fake news on freshly arriving events. With three components in the framework, the first component is a multi-modal feature extractor. Text features are extracted using TextCNN, and a pre-trained VGG-19 is used for visual feature extraction after fine-tuning the hyperparameters. The second component is the fake news classifier, which is built upon the multi-modal feature extraction layers. Classification is performed using a softmax layer for predictions. The third component is an event discriminator that classifies a post into one of the K events. Its application is to remove event-specific features from the posts and capture only invariant features across all events. EANN is capable of classifying fake news incoming from any type of event.

Figure 14: Original vs. tampered image


Figure 15: Face-swapped images

All other image classification tasks that use deep neural networks utilize CNN models, which have been pre-trained on the ImageNet dataset. Jin et al. [55] highlight that to use CNNs for fake news classification tasks, it is required to train the models on a specific fake news dataset. For image credibility analysis of online news on visual data, they collect an auxiliary dataset of fake images from tweets with 328 K rumor images. These images are labeled in terms of their credibility polarity. A collection of 0.6 M images is used to pre-train a convolutional network with architecture similar to AlexNet. The model is trained using iterative transfer learning. This domain transferred learning algorithm provides favorable results for fake news classification.

Qi et al. [56] are the first to use multi-domain information to classify fake and real images proposing MVNN (Multi-domain Visual Neural Network). They expressed that fake news images are constructed of different physical and semantic features than in real news images. They fused the frequency domain and pixel domain feature extraction sub-networks using an attention mechanism to identify real and fake images. Their proposed network consists

of three components: a CNN-based frequency domain sub-network, a pixel domain sub-network built with CNN-RNN to extract semantic features, and a fusion subnetwork. As a whole, their network classifies fake news by using only image features, disregarding linguistics, which is a limitation of the task.

Vishwakarma et al. [57] performed FND by scrapping and authenticating web searches. The work proposes an architecture that enables the verification of text-embedded images. Extraction of text from images is supported by Optical Character Recognition (OCR). The extracted text undergoes Name Entity Recognition steps to obtain named entities mentioned in the text. These strings are used as queries for Google Search to find matching results. The resulting links are categorized into reliable or unreliable. Next, the entities extracted from images are checked against the search result articles' titles, further classifying them into real or fake.

Pasquini et al. [58] also verified the integrity of online news. They have also demonstrated that fake news detection in visual has a dependency on visual forensics. Their focus is to detect news web pages that contain misleading images. Their framework automatically detects related news articles on the internet with similar images of an event. The detection is performed using metadata features present along with the images and by comparing similar features in the set of related images.

Using image forensics techniques for fake news classification, Huh et al. [59] utilized self-learned consistency to detect image manipulation and splicing. It is done by comparing patches of images. Tampered images show low consistency scores between patches. Models are evaluated for splice detection and splice localization on unannotated images. It is an efficient way to detect fake/manipulated images where copy-move forgery, object addition, or removal have been done.

SAME (Sentiment-Aware Multi-modal Embedding) is a deep end-to-end embedding network that exploits users' sentiments to classify fake news. Cui et al. [60] have incorporated users' sentiments with images, profiles, content, and comments into a multi-modal framework. The model intakes post content, image, and user profile and feeds them to a Multi-Layer Perceptron (MLP), pre-trained VGG-19 network, and another MLP, respectively. The three modalities are fused using an adversarial mechanism.

Table 2: Summary of crucial work

Reference	Task	Modality	Technique	Classification	Dataset	Accuracy/F1 Score
[61]	Tweet labeling for FND	Text, Image	Two-level classification (Topic and message-level)	Binary	Twitter (Mediaeval 2015[1])	0.94 (F-score)
[62]	Fake news detection	Text, Image	Utilizing user-profile features	Binary	FakeNewsNet	>0.90 (F-score)
[22]	Fake news detection	Text, Image	BERT, VGG19	Binary	Twitter, Weibo	77.77%, 89.23%
[55]	Fake news detection	Image	Image splice detection, splice localization	Binary	Columbia, Carvalho, Realistic Tampering, In The Wild, Hays and Efros	0.91 (mAP)
[65]	Fake tweet and its user identification	Text, Image	Reverse image search, User analysis, crowdsourcing	Multi-class (fake, legitimate, not sure)	Twitter	
[66]	Fake image classification of Hurricane Sandy	Text, Image	Temporal analysis, Naive Bayes, decision tree	Binary	Twitter	97%
[67]	Image tampering detection, text-image coherence detection	Text, Image	Image forensics	Binary	Mediaeval2016, BuzzFeedNews, CrawlerNews	>75%
[68]	Fake news detection	Text, Image	CNN	Binary	TI-CNN	0.92 (F-score)
[110]	Microblogs news verification	Image	Visual and statistical features	Binary	Sina Weibo	83.6%
[56]	Image verification	Image	Metadata, feature analysis	Binary		72.7%, 88%
[111]	Fake news detection	Text, Image	Bi-LSTM, VGG19	Binary	Twitter, Weibo	74.5%, 82.4%
[72]	Video annotation	Video	DCT, HSV, SURF, AOF, ResNet, GoogLeNet		Twitter	0.40 (F-score)
[112]	Video verification	Video	Contextual Cues	Binary	IVC, FVC	0.9 (F-score)

[97]	Fake news detection	Text, Image	VGGNet, sentiment analysis	Binary	PolitiFact, GossipCop	~75, ~80 (Macro, Micro F1)
[92]	Fake news detection	Image	CNN, Iterative Transfer Learning	Binary	Weibo	77%
[113]	Rumour detection on microblogs	Text, Image	LSTM, att-RNN, VGG19	Binary	Twitter, Weibo	78%, 68%
[93]	Fake news detection	Image	CNN (frequency domain), CNN-RNN (pixel domain), Bi-GRU	Binary	Twitter, Weibo	84.6%
[91]	Fake news detection	Text, Image	Text-CNN, VGG19	Binary	Twitter, Weibo	71.5%, 82.7%
[114]	Fake news detection	Text, Image, Source	CNN	Binary	Twitter	82.47% (F-score)
[115]	Semantic Integrity Assessment	Text, Image	MAE, Bi-DNN, VSM	Binary	MAIM, Flickr30, MS COCO	0.75, 0.89, 0.94 (F-scores)
[100]	Fake News Detection	Image	Montage detection (feature-based approach), SIFT, SURF		COCO, INRIA	>90%
[116]	Fact-Checking on image-claim pair	Text, Image	Cosine Similarity, Embedding similarity	Binary	Snopes, Reuters	80.1%
[117]	Fake news detection	Text, Image	Machine Learning Algorithms	Multi-label	Kaggle Fake News Dataset 2017	85.25%
[70]	News consistency verification	Text, Image, Location, Events	CNN		TamperedNews, News400	94%
[69]	Fake News Detection	Text, Image	Text-CNN, Visual, Similarity features	Binary	PolitiFact, GossipCop	87.4%, 83.8%

Jin et al. [61] performed a classification between real and fake tweets using a two-level classification model: topic level and message level. They proposed that tweets among themselves have a strong relationship in terms of event or topic, and tweets clustered in a topic have similar credibility values. It provided a better result than uni-level classification. Tweets

with similar images (verifying by features such as resolution, image popularity, etc.) were clustered under the same topic. They extracted topic-level features and message-level content features, user features, and other available features and classified them on both levels. Topic-level classification. Results were fused into message-level feature-vector as an extra feature and then trained the classifier. Each tweet was given a separate label instead of labeling each event.

Shu et al. [62] presented a way to utilize user profile features for fake news detection. They extracted and studied explicit and implicit user profile features, also studying which users were most likely to share real and fake news. They have studied users' geolocations, profile images (using CNN, pre-trained VGG16 model), and political bias. PCA was used for dimensionality reduction of profile features. They compared fake news detection performance using UPF (User Profile Features) to multiple approaches, including RST (Rhetorical Structure Theory), LIWC (Linguistic Inquiry and Word Count), RST_UPF, and LIWC_UPF concluding that UPF and UPF-allied techniques provided higher accuracies than others. They also proved that implicit and explicit features, when combined, provided greater results than each being used individually.

Ajao et al. [63] applied hybrid CNN and LSTM-RNN models for text and image classification. LSTM RNN was used to process and classify text sequences. Another model used was LSTM along with dropout regularization 0.2 to remove over-fitting. The third model incorporated a CNN layer after the word-embedding layer of the LSTM model. The models plain vanilla LSTM, LSTM-CNN model, and LSTM with dropout regularization performed in the order of decreasing accuracies 82%, 80%, and 74%, respectively. Under-fitting and lack of sufficient training data account for the low accuracy of the LSTM-drop model. They also showed that hybrid deep learning models' usage provides considerably good accuracy without requiring huge training data.

Singhal et al. [64] have designed a model named SpotFake for detecting multi-modal fake news without any subtasks like finding correlations between textual and visual data. The model consists of a textual feature extractor module that uses BERT (Bidirectional Encoder Representations from Transformers), a visual feature extractor module that uses VGG19, and a fusion module using simple concatenation. The results outperformed state-of-the-art methods

EANN and MVAE and provided higher accuracies equal to 77.77% and 89.23% on Twitter and Weibo, respectively.

Figure 16: Out-of-context image (Mumbai's park image accompanied with Hyde Park's news), (Credit: theconversation.com)

Figure 17: Out-of-context image (A lion taken to vet displayed as a lion been strapped to capture MGM logo), (snopes.com)

Trumper [65] developed a web tool, 'Fake Tweet Buster,' for users to check a tweet's credibility. The user needs to enter a tweet URL, and the application provides a result as fake or legitimate. The tool works on reverse image search (Google Images and TinEye), user analysis, and crowdsourcing. The tool provides the user with matching images, image data (old or new), tweet information, and classification result. It allows the tool-user to enter his opinion about the tweet as fake, legitimate, or unsure. This crowdsourced data can be used in the future to provide value to credit score.

Gupta et al. [66] extracted tweet information and images from Twitter related to Hurricane 'Sandy'. They analyzed that retweets contained many fake images (86%) rather than original tweets. The approach used temporal analysis to study the propagation of fake images over the Twitter network considered a graph with multiple nodes. They created graph networks for followers and retweets of each user studied. For classification, they used user-level features (giving poor results) and tweet level features (providing effective classification), experimenting with two machine learning techniques: Naïve Bayes classifier and J48 Decision Tree classifier, providing 91% and 97% accuracy, respectively.

Lago et al. [67] created a fusion of image forensics algorithms to distinguish between real and fake images. The various approaches used are Image Forensics Methods (Error Level Analysis, Block Artifact Grid Detection, Double Quantization Likelihood Map, Median-filter noise residue inconsistencies detection, JPEG Ghosts, Color Filter Array), Splicebuster, and

with the use of CNNs. They also verified the coherence of online news with their accompanying images. Texts were analyzed using TF-IDF or its higher version, STF-IDF. To make classifications, they chose Random Forest and Logistic Regression.

Yang et al. [68] used convolutional networks for both textual and visual fake news detection. They analyzed latent features and explicit features of text and image sub-branches and fused them using early fusion to classify news.

Huckle and White [69] utilized blockchain and cryptography to trace the origination of fake content. Their focus is to determine where the fake information comes from rather than analyzing its structure and features. Their approach lies in determining the cryptographic hash functions of the text and associated image. The verification is made by matching the hashes of the versions of an image occurring at different places. Same images from different occurrences resulting in different hashes indicate that the image has been altered. This principle forms the base of their fake news detection mechanism.

Krishnan and Chen [70] identified tweets containing fake news using data mining, statistical analysis, reverse image search, and cross verification. They have divided their framework into two components: Core and Website. The Core module fetches tweets, extracts the feature set, performs classification, and returns predictions. The Website module is mainly responsible for crowdsourcing and collecting users' guesses about the post's credibility. It also returns the final decision about the post to the end-user. Classification is performed using the J48 decision tree classifier and Support Vector Machine (SVM). The crowdsourced data is stored in a database for future re-training of classification models and performance improvement.

The NewsVallum approach introduced by Armano et al. [71] uses text-image semantics for fake news detection. It focuses on multi-modality using text and image features for classification with deep neural networks and reinforcement learning. It evaluates the credibility of news spread online on a daily or hourly basis.

Zhou et al. [72] exploited the similarity between text and images to detect if a news article is true. The proposed model comprises three modules where the first one is a multimodal feature-extractor, the second module is a classifier, and the third module determines the cross-

modal similarity between text and image in the post. They have used a text-based one-dimensional convolutional network for text-classification, and image features are extracted using VGG. Both the feature extractors are combined using a concatenation operation. In the third module, the semantic relation between the text and image pair is calculated using cosine-similarity. This helps in identifying if the modalities of a post express the same meaning. A post is classified as fake or real, depending on all of these features.

Chen et al. [73] argue the need for an automatic fake news detector tool for evaluating the integrity of any news online.

Using multi-modal features, Budack et al. [74] measured the consistency between the modalities for fake news verification. The proposed work evaluates the coherence between text and image data in an unsupervised manner. Textual module extracts persons, entities, and locations using Named-Entity Recognition (NER). POS Tagging is applied, and subsequently, embeddings are calculated using fastText. For visual features, the ResNet model is used for verification. Persons, locations, events, and scene context verification is performed in the cross-modal entity verification process. This model applies to real-world news classification.

Parikh et al. [75] performed the task on tweet text and images providing a web application. The UI allows users to upload a screengrab of a tweet from which the model extracts useful information like tweet text, image, username, timestamp, location, etc., and predict the authenticity of a tweet using these features.

It has become easier to create fake scenarios in videos by replacing, removing, or adding objects, adding text, and swapping face in the recent technological eras. These manipulated videos in which a person's face are swapped by another's are termed as 'deep-fakes.' Such videos have created nasty issues of defamation. Forensics based detection methods majorly detect copy-move manipulations or tampered frames. Neural networks have been applied to detect morphed faces in visual data.

Nixon et al. [76] annotated videos as real or fake to verify the news. They analyzed the text of news stories and videos circulating online and used their metadata to fact-check and annotate the videos. In textual analysis, the author checks the stories for correctness, distinctiveness, homogeneity, and completeness and then groups them under clusters. The

videos related to these news stories were then retrieved and annotated based upon fragment information.

Bagade et al. [77] developed a fact-checking web and mobile application 'Kauwa-Kaate' for full-article verification incorporating text, images, and videos. Their proposed system provides a user-friendly interface to query and fact-check information as and when they encounter it. The algorithm scrapes news articles from fact-checked and trusted news sources available on the internet and maintains a repository in the backend. The verification is carried out by matching the query item with news articles in the repository. Devoting a platform entirely for fact-checking, is a very practical method for users to verify fake news.

Using several tweet-based and user-based features, Boididou et al. [78] have introduced a model that predicts tweets as fake or real depending upon the majority vote from individual classifiers that use different features. The approach has displayed a successful classification of news from a variety of events.

An event rumor detection mechanism for Sina Weibo has been designed by Sun et al. [79]. The model uses many features, which are content-based linguistic features, user-based features, and multi-media features. The model is suitable for detecting rumors in the form of text, image, and video.

A new convolutional layer has been proposed by Bayar and Stamm [80] to classify unaltered and manipulated images. Fake images can be of the misleading type where news content and accompanying image are unrelated or of the tampered type where images are forged to create a fake scenario. Detecting forged images can prove highly beneficial to detect fake news and news containing fake images. The new layer can learn manipulation detection features without needing to extract preliminary features. This layer has been incorporated into a CNN architecture to detect multiple manipulations. For model training, ReLU activation and SGD (Stochastic Gradient Descent) is used. CNNs train faster with ReLU. Binary classification classifies images into unaltered vs. tampered images with 99.31% accuracy. Multi-class classification is used to classify images based on types of forgeries used: median filtering, Gaussian blurring additive white Gaussian noise, re-sampling vs. authentic image with 99.10% accuracy.

Table 3: Examples of data manipulation detection techniques

Reference	Task	Modality	Technique	Classification	Dataset	Accuracy/F1 Score
[39]	Deception detection	Text, Video, Audio	3D-CNN, CNN, openSMILE	Binary	Courtroom trials	96.14%
[105]	Tampered video detection	Video, Audio	Speaker inconsistency detection	Binary	VidTIMIT, AMI, GRID	<1% (EER)
[81]	Image repurposing detection	Text, Image, Location	CNN (VGG19), Word2Vec	Binary	MEIR	0.80 (F-score)
[102]	Fake video detection	Video	RNN, CNN (InceptionV3), LSTM	Binary	Deepfake videos, HOHA	>97%
[103]	Forged image and video detection	Image, Video	Capsule Forensics, VGG19	Binary	Deepfake	99.23%
[91]	Face manipulation detection	Video	RNN, CNN (ResNet50, DenseNet)	Binary	Deepfake, Face2Face, FaceSwap	96.9%
[83]	Splice detection	Image	CNN (ResNet50), Illumination Maps	Binary	DSO, DSI, Columbia	96%
[86]	Fake image detection	Image	CNN ensembles	Binary	CelebA, PGGAN	99.99% (AUROC)
[87]	Image translation detection on compressed and uncompressed images	Image	Conventional methods, CNN (DenseNet, InceptionNetV3, XceptionNet)	Binary	CycleGAN	89%
[88]	GANs vs. real image detection	Image	VGG19	Binary	RAISE, Rahmouni	100%
[99]	GANs Image detection	Image	Co-occurrence matrices, CNN	Binary	cycleGAN, starGAN	99.45%, 93.42%
[80]	Image manipulation detection	Image	New convolutional layer	Binary and multi-class	Various camera images	~99.10%
[104]	Face Spoofing Detection	Image	CNN	Binary	CASIA, REPLAY-ATTACK	<5% (HTER)
[118]	Fake image detection	Image	Mixed Adversarial Generators		FantasticReality	0.61 (mAP)

[98]	Fake image detection	Image	Color disparities	Binary	CelebA, LFW, generated images	>90%

Sabir et al. [81] and Jaiswal et al. [82] have detected image repurposing where unaltered images are put together with false metadata in a news item. Pomari et al. [83], Zampoglou et al. [84], and Wu et al. [85] detected fake images by checking if the images or their portions have been sliced.

Pomari et al. [83] did so by making use of illumination inconsistencies in the image. Computer-generated images and videos have gained huge popularity in the present scenario. It has become fairly easy to generate fabricated content through computers. Such content is entirely unreal or mixed with some kind of real-world entities. The result is a fake product that is not reliable at all. Tariq et al. [86], Marra et al. [87], Nguyen et al. [88], Rahmouni et al. [89], and Rezende et al. [90] have identified fake images generated by computer machines. Nguyen et al. [88] did this using Modular CNN. Sabir et al. [91], Zhang et al. [92], Zhou et al. [93], Dang et al. [94] have detected tampered faces in images, which can be utilized in detecting fake news where faces of celebrities have been swapped to create fake scenarios.

Wu et al. [95] have used supervised learning to trace various types of manipulations like copy-move forgery, object removal, splicing, and other unknown tampering in images. Photoshop is a widely used tool used by content creators to modify visual content. Wang et al. [96] have detected variations created in images using Adobe Photoshop. Wu et al. [97] created BusterNet for copy-move forgery detection. It is a type of forgery where an object from an image is removed from its original location and moved to a different place in the same image. Li et al. [98] used RGB color components in the images to detect changes that occurred due to tampering. Nataraj et al. [99] detected GAN-generated images using co-occurrence matrices. These matrices describe the distribution of co-occurring pixel values or colors. Steinebach et al. [100] recognized image montages for fake news detection. In image montaging, two or more images or their parts are arranged together by cutting, overlapping, pasting, etc., to make a composite image.

Korshunov and Marcel [101] and Guera and Delp [102] addressed facial manipulations in videos where the face of a person is replaced by another. They used deep neural networks for the task. Guera and Delp [102] also contributed with a dataset of 300 deep-fake videos

extracted from websites. They classified videos using CNN and LSTM into pristine and deep-fake categories. CNN has been used for feature extraction from video frames, concatenated and propagated to LSTM for analyzing sequences temporally. This architecture allows detecting fake videos as short as 2 seconds of length. Korshunov and Marcel [101] showed that using static frame features correspond to higher accuracies than using audio-visual analysis.

Nguyen et al. [103] identified forgeries like replay attacks, computer-generated images/videos by building a capsule network with CNN layers. Videos are analyzed at frame level, and the probabilities of fake and real of every frame are averaged to generate results for the video.

Yang et al. [104] detected swapped faces in images and videos using CNN. Korshunov and Marcel [105] performed tampered video detection using inconsistencies between audio and video features. Classifiers used were GMM (Gaussian Mixture Model), SVM, MLP, and LSTM. Krishnamurthy et al. [39] performed deception detection over a small dataset of 121 courtroom videos. They used Text CNN for textual analysis, 3D-CNN for videos, and openSMILE with MLP for audio analysis. They also utilized micro-expression features such as smile, laughter, frown as another modality for deception detection. Data fusion techniques used were Concatenation and Hadamard + Concatenation. Wu et al. [106] and Rosas et al. [107] have also proposed deception detection on videos using real-life trial data.

Li et al. [108] detected tampered faces in videos to detect swapped faces of celebrities by detecting eye-blinking features. Eye blinking is detected using the LRCN (Long-term Recurrent Convolutional Networks) model, measuring how much open an eye is measured concerning frame coordinates. Features have been extracted using the VGG16 convolutional layer and propagated to a sequence learning module that uses LSTM-RNN, which can retain memory. Then, the probability is determined of how much the eye is open or closed. Bestagini et al. [109] detected local tampering in video sequences. This was done by finding duplicated frames in videos and cross-correlating them with Spatio-temporal frame regions.

2.5 BENCHMARK DATASETS

Lack of suitable multi-modal datasets have, a lot, hampered the progress in the direction of fake news detection. Deep learning algorithms largely depend on huge amounts of training data, which, being meager, has appeared as a big challenge. The maximum number of fake news detection frameworks built to date have been trained upon data extracted from Twitter, Sina Weibo, or some websites. A few of the other small-sized datasets have been generated for image

Table 4: Benchmark Multi-modal datasets

Dataset	Year	Type	Description	Class	Size	Source
MediaEval[1]	2015	Tweet, Image	Tweets related to 11 events (dev set) and 17 events (test set)	Binary	6,225 real tweets 9,596 fake tweets	Topsy, Twitter API
FakeNewsNet[2]	2018	Text, Image	News Content, Social Context, Dynamic Information, article URLs	Binary	6,000 Fake and 18,000 Real	PolitiFact, GossipCop
Fakeddit[3] [119]	2019	Text, Image	Text, Image, Metadata and Comment	Multi-class	8 Lakh samples	Reddit
Newsbag [120]	2020	Text, Image	News articles and images	Binary	15,000 fake and 2 lakhs real	The Onion, The Wall Street Journal
Newsbag++ [120]	2020	Text, Image	Created by Data Augmentation	Binary	3,89,000 Fake and 2,00,000 Real	The Onion, The Wall Street Journal
FVC[4] [112]	2017	Text, Video	Fake Video Corpus for fake video detection	Binary	55 Fake and 49 Real videos	YouTube

[1] https://github.com/MKLab-ITI/image-verification-corpus/tree/master/mediaeval2015
[2] https://github.com/KaiDMML/FakeNewsNet
[3] https://github.com/entitize/fakeddit
[4] https://github.com/MKLab-ITI/fake-video-corpus

Twitter [113]		Text, Image	Text, Image, Social Context	Binary	6,026 Real and 7,898 Fake	Twitter
Sina Weibo [113]		Text, Image	Data extracted from Chinese online social platform	Binary	4,749 Fake and 4,779 Real articles	Sina Weibo
Politifact [60]		Text, Image	News Content and their corresponding images, Retweet Comments	Binary	432 Fake and 624 Real News	Twitter
GossipCop [60]		Text, Image	News Content and their corresponding images, Retweet Comments	Binary	5,323 fake 16,817 real	Magazines, Newspapers and Social Media
CrawlerNews [67]	2017	Text, Image	News articles and images	Binary	2,500 Images	Google News
Mediaeval 2016[5] [112]	2016	Text, Image	Tweets and images from 53 past events	Binary	17,857 Tweets with 10,628 fake and 7,229 Real	Twitter API
MFN [64]	2018	Text, Image	Event Centric dataset of tweets and corresponding images	Binary	14,000 Tweets and 500 Images 1,154 Real and 1,154 Fake News Articles	Twitter, Snopes and Webhose
TI – CNN[6] [68]	2019	Text, Image	Text, metadata and Image URLs	Binary	20,015 total articles with 8,074 Real and 11,941 Fake	Over 240 Websites
[117]	2019	Text, Image	Metadata, News Articles	Binary	3,568 Fake, 15,915 Real	Multiple Websites

[5] https://github.com/MKLab-ITI/image-verification-corpus/tree/master/mediaeval2016
[6] https://drive.google.com/file/d/0B3e3qZpPtccsMFo5bk9Ib3VCc2c/view

ReCOVery Dataset [121]	2020	Text, Image, Social Information	News Articles, Tweets related to CovID-19.	Binary	2,029 News Articles, 1,40,820 Tweets	Twitter, Multiple Websites
TamperedNews [74]	2020	Text, Image	News articles and images	Binary	72,561 News Articles	BreakingNews Dataset
News400 Dataset [74]		Text, Image	Tweets, Articles	Binary	400	Twitter, Websites
WhatsApp Dataset [8]	2020	Image	Fake images extracted from WhatsApp groups		8,44,000	WhatsApp

analysis, which still is limited in number and not of optimum quality. Video datasets for this task are very rare and contain videos in the count of 100-200. Other video datasets are not completely relevant. There is an urgent demand for good quality multi-modal datasets that would furnish the need of the hour. The advancements in data augmentation or computer-generated data are beginning to contribute towards building datasets. For the time being, we present a piece of tabulated information about available datasets (image, video, and multi-modal) that have been used in the above-reviewed articles for fake news detection and similar tasks (Table 4). We also list out such datasets that contain news article URL or image/video URL. These datasets can be further improved by extracting visual data using web scraping methods.

2.6 PERFORMANCE COMPARISON

In this section, we demonstrate the usage of evaluation metrics utilized by fake news detection tasks and compare their performances based on the most utilized metrics, i.e., accuracy and F1-score. The comparison provided here is irrespective of the dataset but highlights each task's features and methods. We determine how the results have been moving all these years and identify prospective detection methods. Performances are displayed for tasks displaying the results achieved by the experiments on datasets they have used. Visual representations are provided for an easy understanding of how a given model performs when they use a specific set of features.

Evaluation Metrics

This section discusses the various evaluation parameters utilized for fake news classification tasks. We explain the evaluation methods utilized to examine a model's performance using Accuracy, Precision, Recall, F-score, ROC, AUROC, and EER. In figure 18, a confusion matrix is presented that explains the categorization of rightly and wrongly classified items. We refer to the items as news, images, and videos.

CONFUSION MATRIX	*Actually Real*	*Actually Fake*
Predicted Real	TRUE POSITIVE	FALSE POSITIVE
Predicted Fake	FALSE NEGATIVE	TRUE NEGATIVE

Figure 18: Confusion Matrix

True Positive (TP): This includes rightly predicted positive values, i.e., a piece of real news, image or video is classified as real.

True Negative (TN): This includes rightly predicted negative values, i.e., an originally fake item is correctly predicted as fake.

False Positive (FP): This class contains values where a fake item is wrongly predicted as real.

False Negative (FN): This contains real items wrongly predicted as fake by the model.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \qquad True\ Positive\ Rate\ (TPR) = \frac{TP}{TP+FN}$$

$$F1\ Score = \frac{2(Recall \times Precision)}{(Recall+Precision)} \qquad False\ Positive\ Rate\ (FPR) = \frac{FP}{TN+FP}$$

$$Precision = \frac{TP}{TP+FP} \qquad Recall = \frac{TP}{TP+FN}$$

Table 5: Evaluation metrics used in reviewed articles

References	Accuracy	Precision	F1	Recall	AUC	EER	Others
[85], [61], [67], [68], [70], [76], [81]		✓	✓	✓			
[54], [55], [56], [97], [62], [63], [67], [72], [75], [112], [113], [110], [111]	✓	✓	✓	✓			
[80], [53], [102], [116], [96]	✓						
[101]						✓	
[59], [84], [60]							✓
[115]			✓				
[93], [94], [108], [109], [106]					✓		
[89], [74]	✓				✓		
[49]		✓	✓				
[92], [117]	✓	✓	✓	✓	✓		
[95]			✓		✓		
[66]	✓		✓				
[82]	✓		✓		✓		
[100]		✓					

These values allow us to calculate the corresponding accuracy, precision, recall, and F-score. AUC (Area Under Curve) and ROC (Receiver Operating Characteristics), also called AUROC (Area Under Receiver Operating Characteristics), are calculated using TPR and FPR. The ROC curve is plotted with FPR on the x-axis against TPR on the y-axis. AUC is measured as the area under this curve. AUC values range as real values between 0 and 1, with values closest to 1 being good or correct classification while values closest to 0 being the worst with poor classification. EER, defined as Equal Error Rate, calculates the error that occurred in classification. Table 5 identifies the evaluation metrics used by the reviewed tasks.

Result Analysis

Several metrics and parameters have been developed to define the functional performance or, in simpler terms, the algorithms' efficiency on giving the desired output of classification of data from a given dataset. Among them and widely utilized and relied upon

metrics are Accuracy (in percentage), Precision, Recall, and F1 values, among several others such as AUC and EER. Almost all major research related to Fake News detection utilizes one or more among the former set of four metrics (Accuracy, Precision, Recall, and F-Score).

Hence, we bring forth such metric evaluation summary of the most relevant and pivotal experiments conducted for fake news detection. Figure 19demonstrates the results of tasks in terms of F1-scores. We observe that the overall performance stays between 80-95% for methods that use textual and visual features combined. The video classification task, which uses the annotation technique, still has a long way to go. In terms of accuracy (Figure 20), we observe that the range of results is between 70-100%, with an average score of 85%. The majority of fake news classification tasks have relied upon deep neural networks. With the changing time, we also notice an inclination towards forensic algorithms for the same. Trends determine that most of the existing approaches have preferred to use deep learning algorithms due to their efficiency, robust nature, feasibility, and accuracy. Most works have preferred to use more than one feature, i.e., using multi-modal data. Thus, depending on more options and the type of fake information posts can offer. The aim is to consider all parameters that form/alters a user's perception of a piece of information. Convolutional Neural Networks, with maximum usage in the reviewed articles, have displayed eminent classification performance by exploiting the implicit features. They hold the potential to provide better results in future implementations.

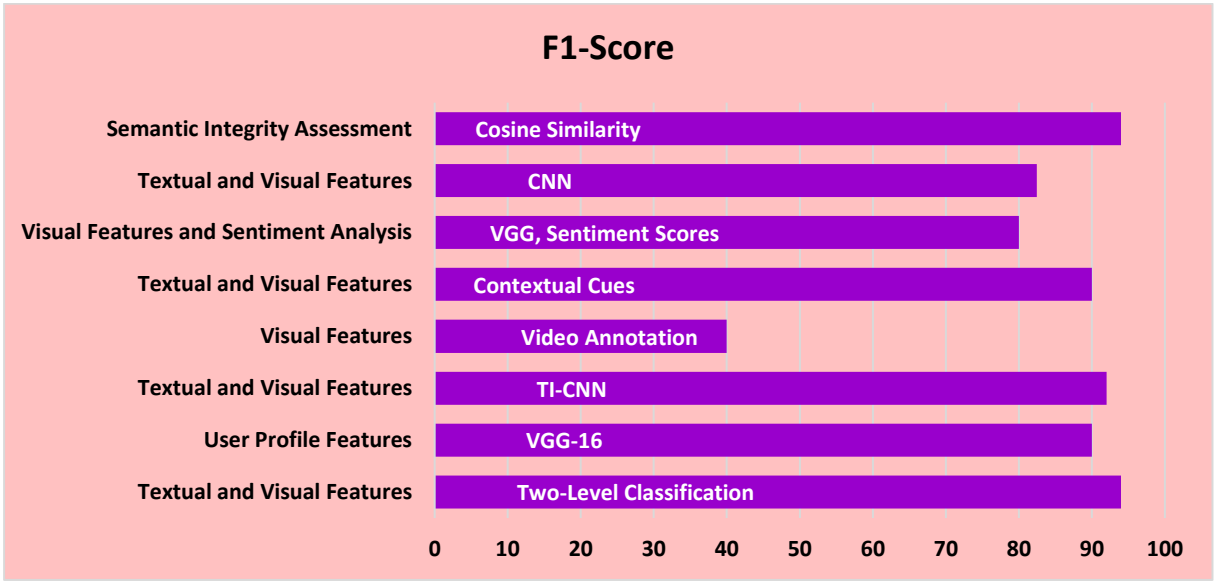


Figure19: Performance comparison based on F1 scores

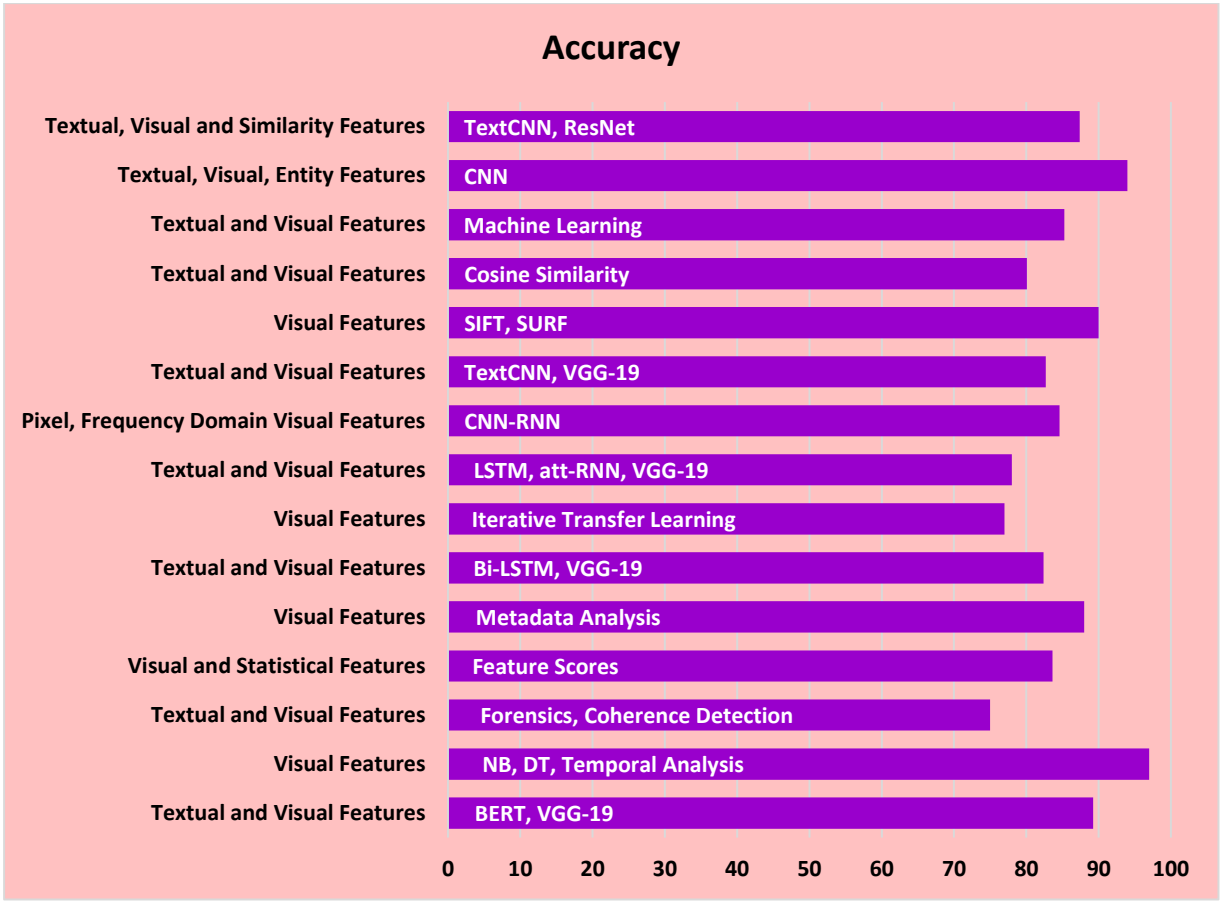


Figure 20: Performance comparison based on accuracy scores

2.7 SUMMARY

Uncontrolled and unauthentic data being over-loaded on the web needs appropriate solutions for the complexities being generated and has become a hard nut to crack. Deep learning algorithms are proving efficient and providing effective solutions with remarkable results. These solutions are to be unearthed from unimaginable horizons and that too within a very precise and limited period as the flow of complexities has reached the verge of parallel solutions. There has been a rapid increase in luring solutions for multimodal fake news detection adopting numerous variant techniques. This survey allows us to conclude that deep learning architectures prove astonishingly capable of fake news detection. They have resulted in high accuracies under the text-domain. Recurrent Neural Networks, LSTMs, GRU, Bi-directional GRU have contributed significantly to text classification. When it comes to visual data, Convolutional Neural Networks form the bigger picture. Survey displays that over 40% of methodologies have incorporated CNNs and their combinations with RNNs or other DNNs in their detection frameworks and served brilliant results. CNNs are taking the lead in computer vision, and allied domains and have become a prospective application for future FND tasks.

Many researchers have identified fake images and videos and tampered regions in them, which we review as supportive tasks that can help classify fake news based on fake visuals. We motivate the readers to combine such tasks with FND modules to perform multimodal FND. By fusing modules performing such tasks on different modalities, optimized performances are assured.

There has been the unavailability of symbolic literature in this domain. The progress along the pathways of multimodal fake news detection has been slow. Researchers are unaware of the advancements so far reached. Existing literature is focused upon fake textual news and its detection mechanisms. We see that accuracy, precision, recall, and F-scores were observed for most of the tasks, whereas some liked to evaluate their models' performances using AUC and ROC. A few other metrics utilized were EER, HTER, TPR, and FPR. Accuracy appears as the most adopted method. Further, owing to the scarcity of multimodal datasets, we regard the obstacles faced in fledged research in the form of not so optimum solutions.

CHAPTER 3

CONVNET FRAMEWORK

3.1 INTRODUCTION

Visual data attracts viewers more quickly than words do. A Human brain captures and rapidly analyzes a news item and often flags it as fake or real by just a glance of its title, image, or a small segment of it, mostly without going through the entire textual content. It does this based on the preexisting knowledge in our conscience. Even if it does go through a whole text, there are very few references and not enough time to check for the authenticity of the content we come across. Various content creators exploit these drawbacks of the human brain and behavior. There is a need for technological state of the art methods to assess the credibility of content, textual or visual, and authenticate it as fake or real. Online media emerged as a platform to share ideas, views, news. With the advancement of mobile devices and the internet, news became easily accessible to people who were either deprived of or uninterested in official news sources such as television and newspapers. The long and seemingly tedious to read texts became easy to understand as images and videos now accompany them. In the same process, it also became challenging to detect the truth in such content.

In the present scenario, online media is losing its charm and credibility as content creators lure users to gain popularity and money using the content they post online. In this process, they do not pay heed to the authenticity of the information, ignore the verification process, and mix up misleading or tampered images or clips with the texts. Content creators focus on posting catchy and attractive content that bags them many likes, comments, and dollars. Sometimes both the text and graphic content are intentionally made erroneous to spread fake news, making the entire content even more unrealistic. Hence there is an urgent need to design and develop a new classification method to assess the credibility of content, textual or visual, and segregate it as fake or real. If textual and visual factors are taken collectively, fake news detection methods have proved to provide higher accuracies than unimodal detection methods. Machine learning and deep learning-based detection mechanisms depend on fake or real news by analyzing the text's features and visual data. Users consuming information play an essential part in stopping the spread of fake content at the root level or circulating to reach a great mass affecting political, social, and economic lives. The algorithms so far used depend upon news data collected from websites and social media platforms, which are later classified into binary (real and fake) or

multiple (ranging according to their severities) labels by crowdsourcing or third-party authenticators.

With the advent of massive data and news content online, the intricacies add up when multiple data forms are available. Despite being beneficial in terms of easy transmission and news consumption, multi-modal data also presents a strenuous task for detecting fake news amongst them. The modalities prevalent on online media include text, image, audio, video, and hyperlinks. With the vast accompaniment of text with visual data, the effectiveness of news rises. A large amount of visual data makes verification difficult as multi-modal data does not guarantee the credibility and attracts more attention than pure text contents. Multi-modal features are expected to be more beneficial in detecting fake news as compared to unimodal features. Few of the excellent quality datasets available for scientific research include binary labeled datasets and multi-label datasets such as Mediaeval, Sina Weibo, PolitiFact, Emergent, and Resized_V2 [1].

Fake news detection challenges include the usage of multi-mal data to classify real and fake news. Present methodologies include fake news detection on textual content [2-4]. Research shows that the incorporation of visual data improves fake news detection. With the rise of multi-modal content on users' posts and news contents, studies involving detection using visual data have rapidly increased.

Previous research [2,5] includes studying image features of visual data like accompanying images, type of image, etc. Other investigations include learning forensic features [6, 7]. Text information is fused by Jin et al. [8] to get better detection using attention mechanism with RNN on image and LSTM on the text and social context to obtain features and perform rumor detection on microblogs. Qi et al. [9] combined Recurrent Neural Networks to detect and interpret real and fake photos semantically. They introduced a novel approach called Multi-domain Visual Neural Network (MVNN). It uses CNN to extract frequency-domain patterns and CNN-RNN to extract pixel domain patterns and fuses using an attention mechanism outperforming state-of-the-art methods by 9.2%. Researchers have provided various forensics tools and techniques to identify image manipulations. Mostly used methods include detecting physical cues within the image.

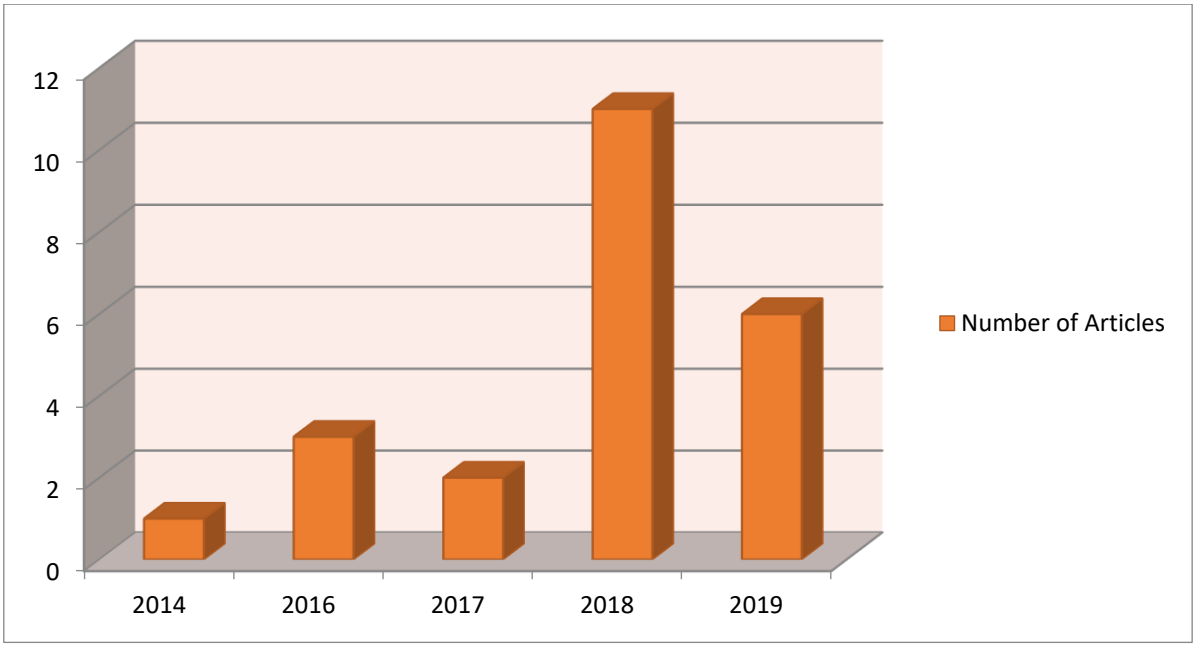


Figure 25: Yearly trend of research works

Recent works have traversed towards deep learning techniques rather than using available prior knowledge about the data. Using labeled training data is specifically advanced for fake news detection. Previous studies focused on linguistic and textual data to study fake news characteristics and semantics of the data. Deep Neural Networks have been utilized to check tweets for temporal-linguistic traits [3]. Attention mechanisms have also been used with RNNs for fusion [10]. Liu and Wu [11] modeled the classification with a combination of CNNs and RNNs. Less focus has been given to the credibility of multi-modal data on the web. Text and images can be well represented using deep neural networks. Jin et al. and Wang et al. [8, 12] applied it to fake news detection.

To overcome the limitation of learning shared representation of multi-modal data, Khattar et al. [13] proposed a Multi-modal Variational AutoEncoder. It is coupled with binary classifier features of text and image modalities with three components in the model, an encoder, decoder, and a fake news detection module. The model leverages state-of-the-art techniques with ~6% accuracy. Ajao et al. [14] used a hybrid of CNNs and LSTM-RNNs to identify fake news-related features without prior knowledge, achieving 82% accuracy. Jindal et al. [15] presented two novel datasets containing fake news text and image, using data augmentation to increase fake news data. Singhal et al. [16] perform fake news detection by introducing the SpotFake framework that exploits textual and visual features of news posts without considering subtasks such as event discriminator and modality correlations. The model increases the accuracy from previous approaches by 3.27% and 6.83% on Twitter and Sina Weibo datasets. TI-CNN has

been proposed for fake news detection by Yang et al. [1] using Convolutional Networks on both textual and visual data. They have incorporated both explicit and latent features extracted for both the modalities using CNN layers.

A new challenge emerged to detect fake or computer-generated images with technological advancement in Generative Adversarial Networks. GANs pose a threat by allowing the creation of fake images and manipulations in existing images. Marra et al. [17] studied the performance of existing detectors that use conventional and deep learning methods, concluding higher efficiencies by deep learning detectors with 89% accuracy. They compared the performances of traditional and deep learning image forgery detectors on a dataset of 36302 images under compression and without compression, concluding that high accuracies are obtained on compressed data using deep networks like XceptionNet, InceptionV3, and DenseNet. In recent years, the yearly trend of published articles using deep networks for credibility analysis is represented in Figure 3. Figure 4 offers the percentage of fine-tuned CNN models in similar tasks.

By extracting event-invariant features, proposing event adversarial neural networks, Wang et al. [17] performed fake news detection on newly arrived events. Three tasks are completed, namely feature extraction, detection, and event discrimination. The study is conducted by ignoring features that are event-specific and considering just the shared features. It provides accuracies of 71.5% and 82.7% on Twitter and Weibo, respectively. Sabir et al. [18] detected image repurposing, i.e., manipulations in image meta-data on a self-proposed MEIR dataset that consists of real-world Flickr data. It proposes a multi-modal deep learning method that utilizes metadata and image information to identify modifications.

Pomari et al. [19] came up using CNNs and illumination maps in images to detect splicing in fake images with a colossal accuracy of more than 96%. Another approach used diverse modalities, including text, image, and source, to detect hoaxes [20]. Bayar and Stamm [21] developed a new convolutional layer that learns features from training data suppressing image features and highlighting manipulation features. This new approach can detect image manipulations with an accuracy of 99.10%. Lago et al. [22] performed the task using image forensics algorithms to see tampered images and a verification mechanism to check if the images are rightly mapped to textual news. In 2019, Cui et al. [23], a detection framework named SAME, exploits user comments and latent sentiments and uses an adversarial

mechanism. Volkova et al. [24] performed a qualitative and quantitative analysis of fake news classification models, proposing a qualitative analysis tool ERRFILTER. Modalities analyzed are text, lexical and image inputs, and their combinations.

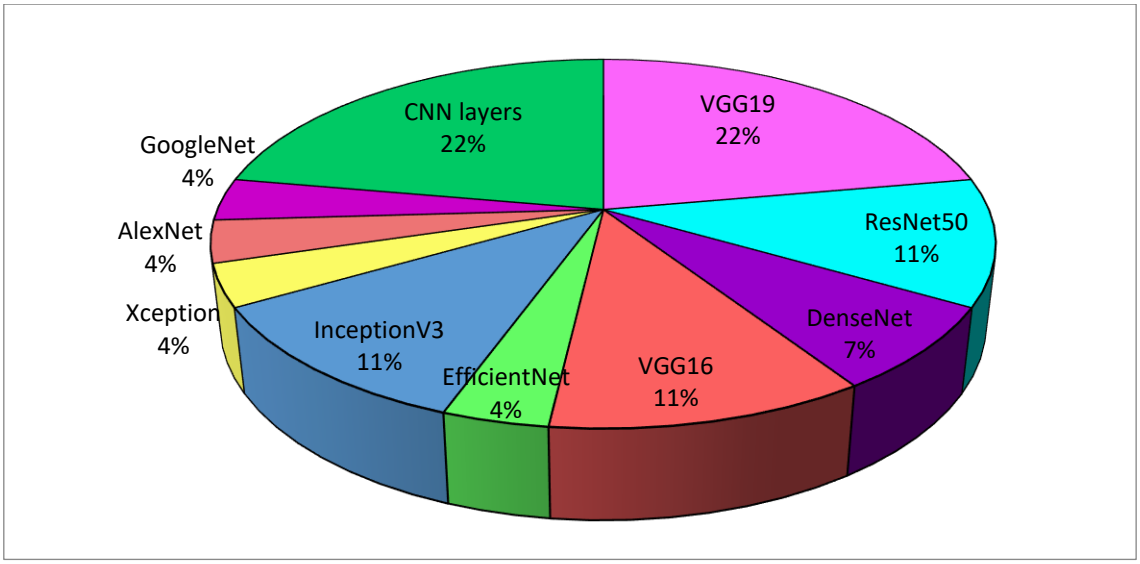


Figure 62: CNN architectures used in previous research

In image classification, Tariq et al. [25] detected fake face images generated by humans and machines using CNN-based models including VGG16, VGG19, ResNet, DenseNet, NASNet, XceptionNet, ShallowNet, and their ensembles. These neural networks detected GANs and human-generated fake face images without using their metadata. The highest accuracies on various image sizes were obtained with Ensemble ShallowNet (V1& V3).

Sabir, Cheng, et al. [26] performed the detection in manipulations of faces in videos using recurrent convolutional models. These models have proved beneficial in utilizing temporal information in still images to detect tampered images improving the existing accuracies by up to 4.55%. Fake video detection has been performed by Guera and Delp [27], using a convolutional LSTM model on a large dataset of deep fake videos in which face swaps have been done. Papadopoulou et al. [28] verified real-time, user-generated online videos, YouTube videos taking their context into account. The information exploited includes video comments for textual data and metadata like video description, likes, dislikes, and uploader information.

Table 6: ConvNet Architectures for credibility analysis of different data modalities

References	Modality	Task	Network	Model
[13]	Text, Image	Fake News Detection using MVAE	RNN, CNN	Bi-LSTM, VGG19
[14]	Text, Image	Fake News Detection on Twitter	Hybrid CNN, RNN	LSTM, CNN
[26]	Video	Face Manipulation Detection	RNN, CNN	ResNet50, DenseNet
[16]	Text, Image	Fake News Detection	RNN, CNN	BERT, VGG19
[1]	Text, Image	Fake News Detection	CNN	Bi-LSTM, CNN
[17]	Image	GAN-generated Fake Image Detection	CNN	DenseNet, InceptionV3, Xception
[12]	Text, Image	Fake News Detection	EANN	Text-CNN, VGG19
[18]	Image	Image Repurposing Detection	CNN	VGG19
[27]	Video	Fake Video Detection	RNN, CNN	LSTM, InceptionV3
[19]	Image	Image Splice Detection	CNN	ResNet50
[20]	Text, Image, Source	Hoax Detection	CNN	Deep CNN [1]
[21]	Image	Image Manipulation Detection	CNN	Proposed CNN
[22]	Text, Image	Image Trustworthiness Assessment in Online News	CNN	
[23]	Text, Image	Fake News Detection	CNN	LSTM, VGG16
[25]	Image	Classifying Computer-generated and Photographic Images	Modular CNN	VGG19
[8]	Text, Image, Video	Rumor Detection on Microblogs	Att-RNN	LSTM, VGG19
[29]	Image	Tampered Face Detection	Two Stream Neural Networks	GoogleNet, InceptionV3
[30]	Image	Classifying Computer Graphics and Natural Images	CNN	MLP

We propose that online social media images consist of three features: latent features, explicit features, and contextual features. Latent features are extracted using layers of convolutions. Deep convolutional networks are capable of learning kernel values that are utilized to extract latent features. According to Yang et al. [1], explicit features are hand-crafted features such as the resolution of an image and the number of faces in the picture. Apart from these two intrinsic features, contextual features are based on semantic relationships between the text and the image. We have executed convolutional neural networks for text and image classifications. CNNs provide an advantage to extract features directly from raw input without any pre-processing

required. CNNs reduce input data on various layers such that only required information is preserved and worked upon to make essential predictions. In this work, we propose a novel fake news detection framework. It is based on two-stream convolutional neural networks for text and image input streams. This novel architecture consists of individual text and image classification modules, which are fused at a later stage post-training of convolutional models. The experiments performed resulted in ~3-6% higher scores than the established state-of-the-art methods. The proposed architecture is capable of detecting fake news based on both textual and visual information. The usage of Text-CNN increases the overall efficiency of the architecture. Simultaneously, the combination of Image-CNN has resulted in an additive accuracy for the detection task. The use of convolutional models that we propose with introduced Text-CNN and Image-CNN models outperform the existing state-of-the-art.

The contributions of this work include:

- Web scraping, creating clean-image datasets from two previously available datasets that contained news URLs.
- We have proposed a new Coupled ConvNet architecture that constitutes proposed Text-CNN and Image-CNN modules for multi-modal fake news detection.
- We have implemented CNN models on TI-CNN, Emergent, and MICC-F220 dataset on textual and visual data.
- We have performed a comparative analysis of various CNN models' efficiencies on real-world datasets for fake news detection.
- We have analyzed the performance of deep learning on latent textual and visual features for fake news detection.
- We have provided new deep learning pathways to better fake news detection.

3.2 PROPOSED METHODOLOGY

This section elaborates on the architectures of the classification models utilized in this task. Proposed Coupled ConvNet is composed of Text-CNN module for textual fake news classification and Image-CNN module for visual fake news classification. We pre-process input data at their earlier stages in both modules and feed them to convolutional neural networks. This section explains the architectures and mathematical background of Text-CNN and other CNN models utilized in this work. Table 1 summarizes different Neural Network architectures used for the credibility analysis of data in various modalities.

Text-CNN

CNNs are widely used for visual tasks. For image classification, pixel information extracted from images is propagated as pixel values to consequent convolutional layers. Words are needed to be processed to make them understandable by a machine. A computing machine treats visual and textual data in the same manner as numeric data. The idea is to serve the machines with text in numeric data in the same way visible data is treated using pixel values. This task is performed by embedding words into vectors. Figure 5 details the various layers of text-CNN architecture.

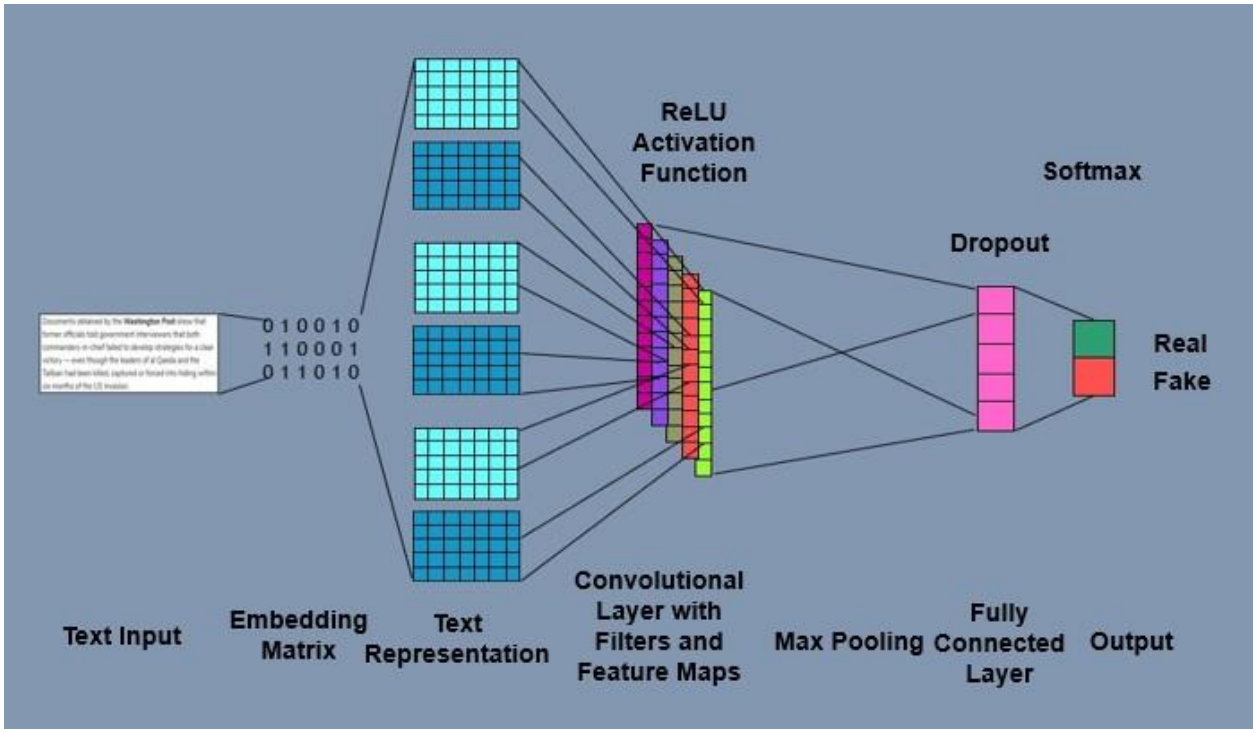


Figure 27: Text-CNN architecture

A fixed vector can thus represent each word in the sentence. These embedded vectors are then propagated through convolutional layers in the same way image data moves through the deep network. The consequent layers are of the same structure incorporating max-pooling, padding, activation function, fully connected layers, and dropout. It is mathematically represented in the form of a k-dimensional vector as $x_i \in R^k$, where x_i is the i^{th} word in a sentence.

Then, $x_{1:n} = x_1 \oplus x_2 \oplus \ldots \oplus x_n$, ($x_{1:n}$ is a sentence of length n) and $\oplus$ represents concatenation. The series of words $x_i, x_{i+1}, \ldots, x_{i+j}$ are concatenated as $x_{i:i+j}$. If h is the number of words, a filter w that is applied to the text generates a feature c_i from a word window

$x_{i:i+h-1}$ where filter $w \in R^{hk}$ and $c_i = f(w \cdot x_{i:i+h-1} + b)$ given b as a bias term and f, a non-linear function. The filter w is applied to every word window, producing a feature map $c = [c_1, c_2, \ldots, c_{n-h+1}]$ where $c \in k^{n-h+1}$. A max-pooling layer is applied next. It extracts the feature with maximum value in the feature map c which is expressed as $\hat{c} = \max\{c\}$. These features with maximum values are propagated further to fully connected layers passing them to a softmax layer for classification.

Image-CNN

The fine-tuned CNN architectures provide good accuracies when it comes to extracting hidden image features and patterns. We implemented eight different CNN architectures AlexNet, Xception, VGG16, VGG19, ResNet50, MobileNetV2, InceptionV3, and DenseNet, for visual fake news detection. The designs of all fine-tuned image CNN models used in this work are described in the following section and represented in Figure 6.

AlexNet: AlexNet is a Convolutional Neural Network designed by Alex Krizhevsky in 2012 and won the ILSVRC challenge. The model displayed that depth in the network is necessary for efficient applications. Depth in the model contributed to providing high performance and became computationally costly, sufficed by using multiple GPUs. AlexNet architecture consists of 8 layers. The first five are convolutional layers, with each layer optionally being followed by a pooling layer and the last three layers are fully connected layers. The model prefers the ReLU activation function, owing to its advantage in training time over tanh or sigmoid functions. Overfitting encountered in AlexNet was reduced by data augmentation and using Dropouts that turn off neurons with a specified probability of 0.5.

VGG: Visual Geometry Group (VGG) won the ILSVRC 2014 competition. The group members, Karen Simonyan and Andrew Zisserman, experimenting with multiple numbers of layers in the deep network, released two versions of their model, VGG16 and VGG19, with 16 and 19 deep network layers each. They displayed that deeper networks with a more significant number of layers result in higher accuracy for image classification tasks. They replaced large kernel-sized filters of sizes 11×11 and 5×5 with smaller filters of size 3×3. Three fully-connected layers follow the convolutional layers following a softmax layer. ReLU is used as the non-linear activation function for hidden layers. The number of channels increases with a twice-factor from 64 in the first layer to 512 in the last layer. The increased depth makes VGG a network slower to train.

ResNet: Residual Neural Network is a network simplified by skipping layers introduced by Kaiming He in 2015. ResNet makes double and triple layers skips jumping across the network. This network makes training more comfortable and faster and reduces the vanishing gradient problem as there is a lesser number of layers in the network. It uses the ReLU activation function and Batch Normalization. Activations are reused from a previous layer until the current layer learns the weights.

Layers are indexed as $l-2$ to l for single skips in backward propagation and as l to $l+2$ for forward propagation. Given $k-1$ as the skip number, this can be generalized as $l-k$ for a backward skip and $l+k$ for a forward skip. A residual network building block with residual function $F(x)$ can be defined by the equations:

For equal dimensions of x and F, $$y = F(x, \{W_i\}) + x \quad (1)$$

and

For unequal dimensions, $$y = F(x, \{W_i\}) + W_s x \quad (2)$$

Here x is the input vector, and y is the output vector, $F(x, \{W_i\})$ is the residual mapping and W_s is a linear projection used for mapping dimensions.

Inception V3: The inception V3 model by Google for image classification was presented in ILSVRC 2015, providing a low error rate due to a 42-layer deep network. This model uses the factorization method to factorize a 5×5 convolution into two 3×3 convolutions. It reduces the parameters by 28%. Similarly, a set of one 1×3 and one 3×1 convolution can be replaced by a 3×3 convolution. The auxiliary loss tower in Inception V1 is used only on the last 17×17 layer as a regularizer in Inception V3. Inception V3 is observed to be much efficient than VGGNet in terms of computation cost.

Xception: Xception stands for "Extreme Inception," Its architecture is entirely based on depthwise separable convolutional layers. Its architecture consists of 36 convolutional layers (as 14 modules) followed by fully connected layers and a logistic regression layer. Except for the first and last modules, all convolutional layers have residual connections. The weight decay rate or L2 regularization of the Inception V3 model was improved to $1e-5$, and the dropout layer used a probability of 0.5. The model does not incorporate the 'Auxiliary loss tower' that is optionally used in Inception V3 architecture.

DenseNet: DenseNets, introduced in 2018, are residual networks with various parallel skips. Each layer in a DenseNet is connected in a feed-forward manner to every other layer. The expression gives the total number of direct connections between the layers $\frac{L(L+1)}{2}$, where L is the number of layers. DenseNets do not require the learning of repeated feature maps and require a lesser number of parameters. They perform concatenation of feature maps instead of sum. Its equation can be stated as:

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \tag{3}$$

Here x_l is the output of l^{th} layer and H_l is a non-linear transformation.

MobileNet V2: MobileNet V2, a type of CNN, was specially designed in 2019 for mobile devices based on inverted residual connections and bottleneck light-weight depthwise separable convolution layers. The first layer of MobileNet V2 is a convolutional layer with 32 filters. Nineteen residual bottleneck layers follow it. The kernel size used is 3×3, and the non-linear activation function used is ReLU6. The residual layers are used to make the model memory efficient. The bottleneck block operator used can be expressed as:

$$F(x) = \sum_{i=1}^{t} (A_i \circ N \circ B_i)(x) \tag{4}$$

where, A_i is a linear transformation, N is a non-linear transformation and B_i is a linear transformation to the output domain.

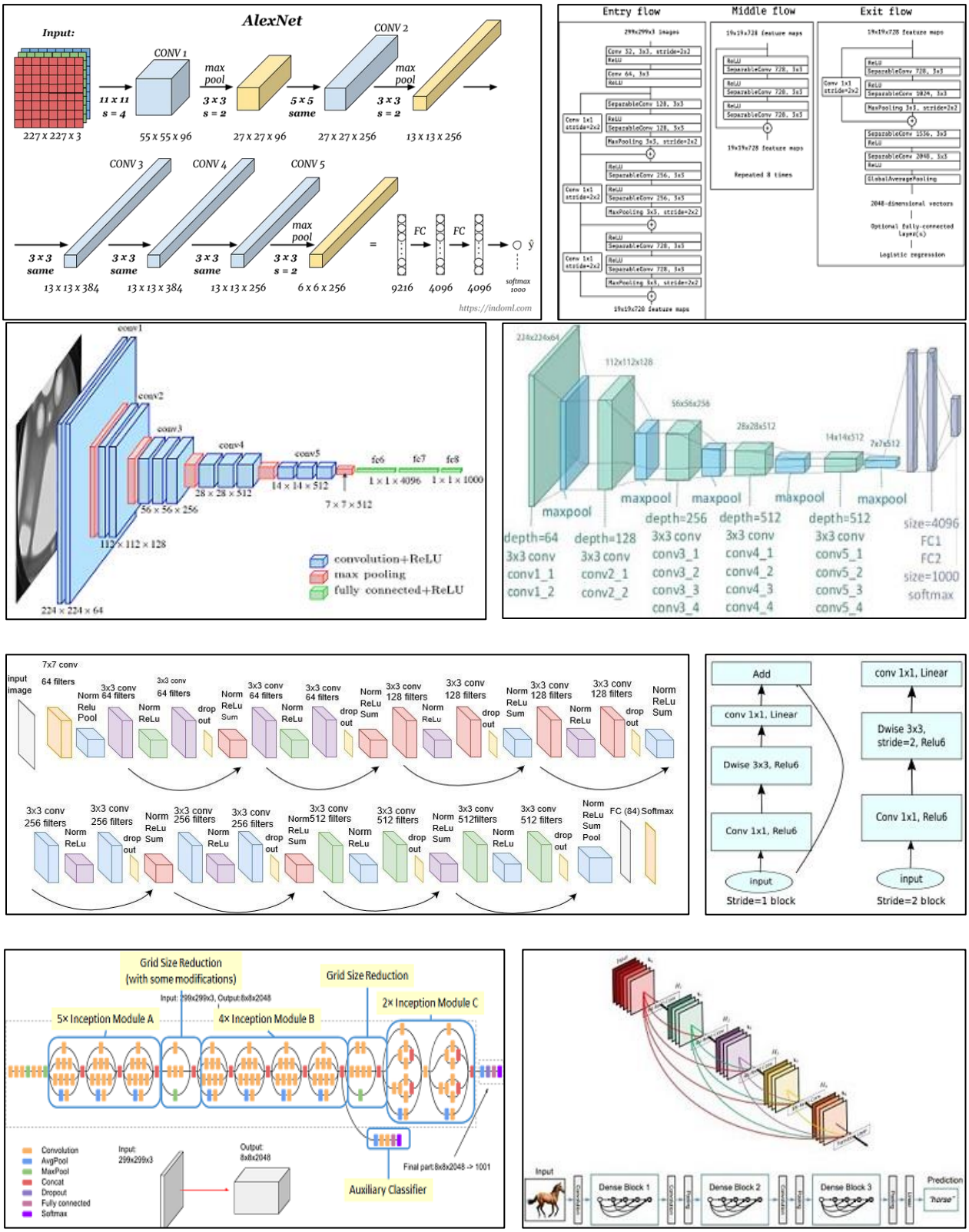


Figure 24: CNN model architectures: (a) AlexNet, (b) Xception, (c) VGG16, (d) VGG19, (e) ResNet50, (f) MobileNetV2, (g) InceptionV3, (h) DenseNet

Proposed Coupled ConvNet

The proposed approach to fake news detection extends the utilization of convolutional neural networks to a broader scale to automate fraudulent content detection on the web. Most of the existing literature is flooded with singular modality tasks where one of the present features are exploited. Most of the approaches are based on machine learning algorithms, while others use deep learning such as GRU, LSTM, Bi-LSTM, and other RNNs for text classification. We leverage this task by introducing a new text classification model using a convolutional neural network. With the onset of using visual features, pre-trained CNN networks are in wide use. The proposed image classification model is based on the usage of a pre-trained model with fine-tuning. Fake news detection tasks can be combined based on data modalities. Hence, the Coupled ConvNet introduced in this work is a hybrid two-stream convolutional architecture (based on text stream and image stream) is proposed, which are then combined using a late fusion technique. The architecture comprises of two streams (modules): Text Module (for textual classification) and Image Module (for visual classification). The architectures of these modules are explained in sections 4.1 and 4.2. The combination mechanism used in the proposed Coupled ConvNet is provided in section 4.3. The series of operations performed in both the modules is depicted in figure 7. Figure 8 represents the proposed Coupled ConvNet architecture.

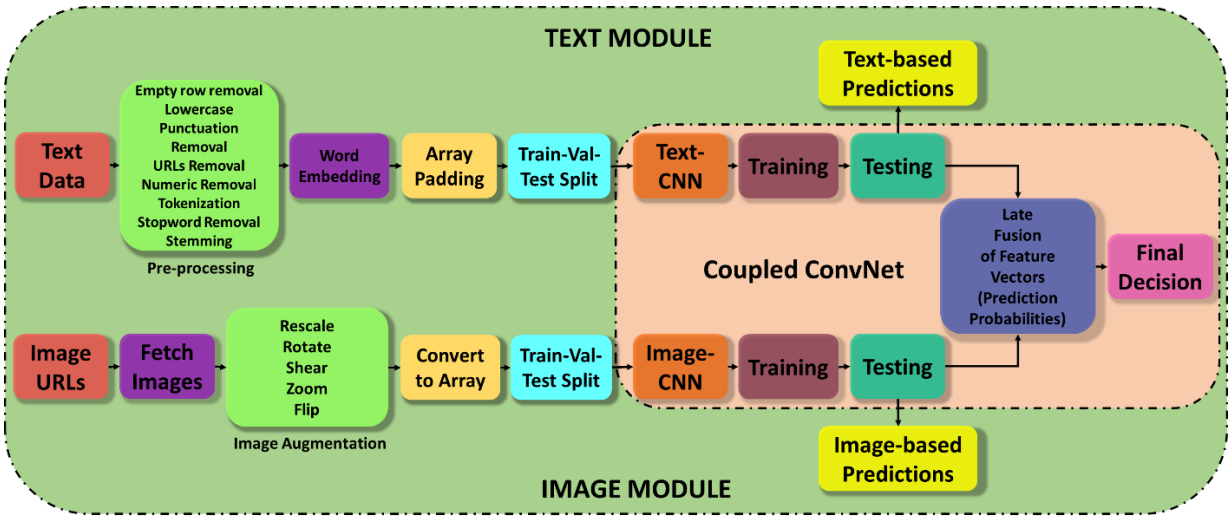


Figure 25: Sequence of operations performed

Text Module

A raw text dataset undergoes several refinements and analysis procedures before being affirmed for its realness. The first of those processes is pre-processing the text information. Then the word embeddings are generated for the textual content. Upon completion of this step,

the embedded vectors are fed to a one-dimensional convolutional model. We then utilize the CNN model on textual data by applying convolutions on text vectors. A series of layers of convolution and pooling are generated to analyze the data features. Finally, all these layers are conjectured to provide a binary output of the data's information's authenticity. The results are obtained after training the data under multiple iterations of the proposed Text-CNN model.

We use only the 'title,' 'text,' and 'label' columns from the versatile information present in the datasets. Textual pre-processing involves the following steps: lowercase conversion, punctuation removal, URL removal, numeric value removal, data tokenization, stop-word removal, and stemming/lemmatization. In the next step, we perform Array Padding. Padding is done by calculating the maximum length from the most extended news item present in the array data. The text, which is shorter in length than the full content length, is padded with zeroes. The data is further split into the train, test, and validation sets. This processed data is now encoded, and text and title inputs are embedded using Glove embeddings. These embeddings are added next to the 1-D input layer. We then feed this data to the proposed CNN model. The proposed text classification model consists of three one-dimensional convolutional layers with ReLU activation function, each followed by a max-pooling layer. Subsequent layers are fully connected Dense and Dropout layers. After experimentation with different dropout values ranging between 0.2 to 0.8, the best results were portrayed by setting the value to 0.4 for both the dropout layers. A binary Sigmoid classifier is deployed to generate the predictions.

Image Module

CNNs have shown considerable performance for various image classification tasks. They identify latent features without demanding any extra information. These latent features are present inside an image and are described as resolution, objects, pixel parameters, size of an image, etc. When the image data under examination is combined with other modalities such as text, it classifies real and fake news. For Image Analysis, the available image datasets are created as explained. The datasets consist of URLs of news pages. We use these URLs present in the database to scrape URLs of images present in those news pages, using BeautifulSoup. We download and then zip the fake and real photos from those newly obtained URLs into separate directories to our local access. These image URLs are also added to the datasets corresponding to their respective news. Data folders are uploaded to Google Drive, and the drive is mount to Google Colab. We use the split-folders module to divide the dataset into train, test, and validation sets with 80%, 10%, and 10% fake and real images, respectively, for TI-

CNN and EMERGENT datasets. MICC-F220 dataset is split into 60%, 20%, and 20% for training, validation, and testing sets, respectively. A different proportion is used for MICC-F220. This difference in splitting ratios consists of 220 images, with 110 real and 110 fake images. Splitting this dataset into 8:1:1 leads to a minimal number of images in the validation and test sets. It creates a bias in the classification results. To avoid this bias and generate normalized results, this dataset has been split in a proportion that keeps a good number of images for validation and testing. After this, we perform Image Augmentation using ImageDataGenerator. Operations performed during augmentation include rescaling, rotation, shear, zoom, and flipping of images, which improves the quality of the datasets for usage. Image data is then fed to various mentioned CNN models for classification. The CNN training sequence is similar to that of the text convolution sequence except that in this case, two-dimensional convolutions are performed on visual (image) data. We feed visual data to various CNN models separately. The list of multiple models experimented with our data includes AlexNet, ResNet50, MobileNet, DenseNet, XceptionNet, InceptionV3, VGG16, and VGG19 [31-34]. Accuracy is determined after training the models for a specified number of epochs, and the result trends for training, test, and validation are observed.

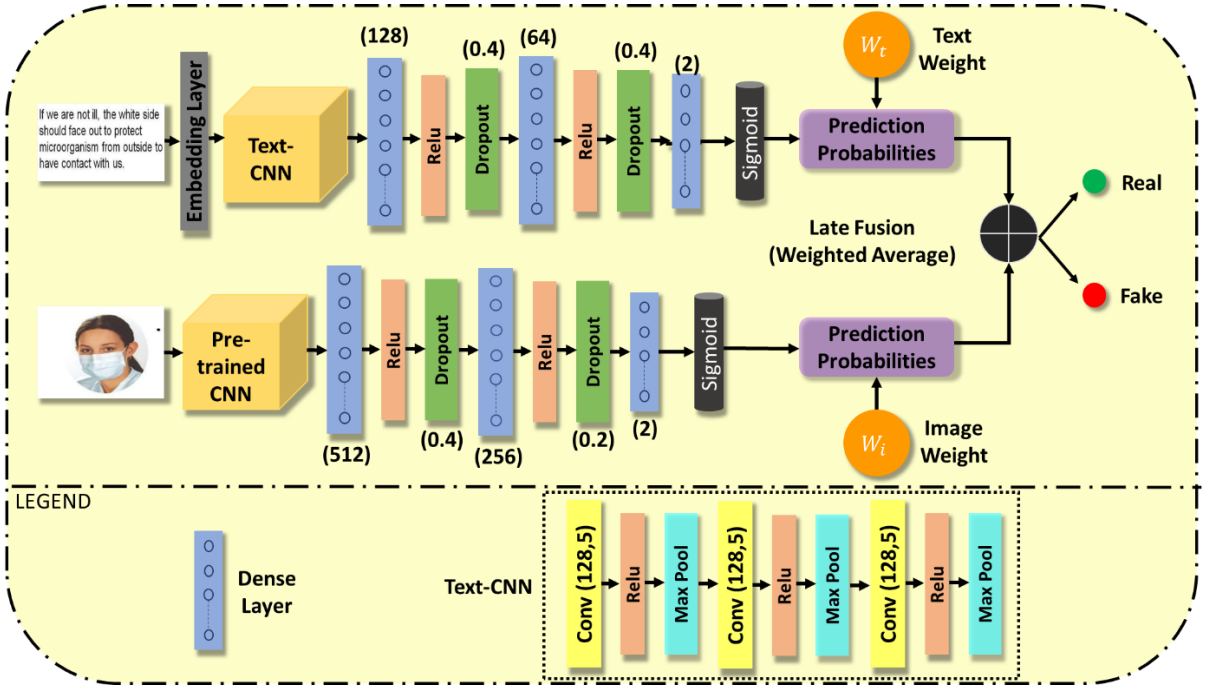


Figure 26: Proposed Coupled ConvNet Architecture

The proposed Image-CNN module uses one of the above-mentioned pre-trained models for each experiment. After adding a pre-trained model, a Dense layer of shape 512 with ReLU activation is added. Next, a Dropout layer of probability 0.4 is used. Another dense layer with shape 256 is used next. Subsequent layers incorporate a dropout layer of value 0.2 and a binary

classification layer with a sigmoid activation function. The dense and dropout values are chosen decreasingly to avoid the immediate transition to the final classification layer. It allows the input to travel smoothly through the fully connected layers rather than directly jumping to the last layer. As observed during the experiments, using two dropout layers of value 0.4 and 0.2 reduces overfitting considerably and reduces the loss during the training phase, thereby increasing accuracy.

Text-Image Fusion

Post-implementation of text and image classification modules separately, this segment fuses the outputs obtained. Prediction probabilities from both the modules are forwarded to a late fusion operation. Late fusion, a scalable and straightforward method, combines the features from multiple streams after the training phase. The decision vectors from each stream are combined using a suitable combinatorial operation. The proposed method uses a weighted fusion approach in which each modality is assigned a weight that determines the contribution of that modality in the final classification decision. Weights are chosen in a way such that maximum classification accuracy is obtained. For a fusion function $f : P_t, P_i \rightarrow P_c$ where P_t *and* P_i are two different sets of prediction probabilities that denote the decisions of each stream, the combined probabilities indicated by P_c gives the output decisions after late fusion. P_c is calculated by adding the products of text and image prediction probabilities with their assigned weights W_t (text-weight) and W_i (image-weight). It is expressed as:

$$P_c = P_t * W_t + P_i * W_i \quad (5)$$

Choice of weights is made by experimenting with all possible combinations, varying the weight values between 0.1 to 0.9, with a difference of 0.1 unit. Text and image weights vary inversely. The variety of probabilities that produce the best result are used for each experiment. These weights have been described in table 2 in section 5.2.

3.3 EXPERIMENTAL RESULT ANALYSIS

Datasets and Preprocessing

TI-CNN: With the availability of only a few good quality multi-modal datasets, we utilize the already collected dataset that is available online[7], used by Yang et al. [1] for a similar

[7] https://drive.google.com/file/d/0B3e3qZpPtccsMFo5bk9Ib3VCc2c/view

fake news classification task. This dataset contains 20,015 news items from websites, with 11,941 items being fake and 8,074 being real. The dataset is rich in terms of the wide range of details that it covers. We use all of these news items for the Text-CNN module using their title, text, and label information. For the Image-CNN module, we use image URLs obtained from the dataset in the 'main_img_url' column to scrape images from the web. The total number of images extracted from TICNN is 5733, constituting 2612 real news images and 3121 fake news images. The remaining URLs redirected to corrupted web pages or pages removed left us with an image dataset of a smaller size than their corresponding text items. TI-CNN dataset is used for experimentation in both Text-CNN and Image-CNN modules and later in the proposed Coupled ConvNet architecture.

EMERGENT: Another dataset experimented with is the EMERGENT (FNC) dataset created by Ferreira et al. [35], consisting of a total of 300 claims and 2595 associated articles. We polish this dataset by discarding duplicate news items and removing blank spaces. For the Image-CNN module, we use post URLs to extract image URLs and then scrape images from EMERGENT datasets' web-pages that led to a clean dataset of 1338 fake and 791 real images. We have made both of these image datasets publicly available. This dataset is also used in both the proposed individual modules and then in the proposed Coupled ConvNet architecture.

MICC-F220: Further, we used the MICC-F220 dataset by Amerini et al. [36] that consists of only real and tampered images, without any other form of data present. We use it with CNN models to identify whether an image is tampered with or original, in short, fake or real. Due to the lack of textual information, this dataset is solely employed in the proposed Image-CNN module. It is used to compare the efficiencies of utilized pre-trained CNN models within the proposed architecture.

Experimental Settings

All experiments have been performed on Google Colab that provides up to 13.53 GB of RAM. It also allocated us 12 GB NVIDIA Tesla K80 GPU hardware accelerator and python version 3. In Text-CNN, we employed RegexTokenizer to extract tokens from news titles and news texts. To reduce the words into their root forms, we used Porter Stemmer and WordNet Lemmatizer. We have utilized Glove representations for word embeddings used in Text-CNN. We have also applied one-dimensional convolutions on title and text and concatenated their layers. We used 0.4 and 0.8 as subsequent dropout values in the experiments. We used a batch size of 64 and have trained the model upon running for 250 epochs. For Image-CNN, we take

the image input in size 224*224. Upon setting the dropout value to 0.2, the experiments exhibited a considerable increase in training accuracy. We have used Adam optimizer for all the given models. The batch size is set to 64 instances. The value of batch-size affects the training time of the model. The aim is to maximize the performance of classification models and minimize computation time. Choosing a batch-size less than 64 resulted in higher training time, which made the process slower. Whereas Google Colab did not accommodate a value greater than 64. Therefore, 64 is the perfect fit and is used as the batch size for both text and image modules. We have used binary cross-entropy loss for classifying the item into two categories: real and fake. In the combinatorial phase, weights for text and image features that provided the best classification accuracies were recorded and are as follows:

Table 7: Fusion Weights that provided maximum classification accuracies

Model	TI-CNN		EMERGENT	
	Text	Image	Text	Image
ResNet50	0.5	0.5	0.8	0.2
VGG16	0.5	0.5	0.5	0.5
VGG19	0.7	0.3	0.6	0.4
InceptionV3	0.8	0.2	0.7	0.3
DenseNet	0.5	0.5	0.8	0.2
Xception	0.5	0.5	0.5	0.5
AlexNet	0.6	0.4	0.7	0.3
MobileNet	0.5	0.5	0.5	0.5

Results

This section presents the performance comparisons of all models used in our work for fake news classification on each of the three datasets. The scores are presented as accuracy, precision, recall, and F1-scores.

The comparison values of the Text-CNN module on two datasets, TI-CNN and EMERGENT indicate that CNNs exhibited an outstanding performance for classifying text-based fake news with 96.26% accuracy on TI-CNN and 93.56% accuracy on EMERGENT. Better scores were obtained on the TI-CNN dataset when compared to EMERGENT in all Text-CNN performance scores. It accounts for the larger size of TI-CNN data. More data aids in better training and hence produces better results. We portray performance comparison values for eight Image-CNN modules on TI-CNN and EMERGENT. VGG16 and VGG19 performed the best with 82.72% and 81.04% scores, respectively, on the TI-CNN dataset, followed by ResNet50 and MobileNet with 77.54% 73.37% accuracy, respectively. Other Image-CNN

models scored below 63% accuracy on the TI-CNN dataset. The top four in terms of precision and F1 score were in the same order as the accuracy on the TI-CNN dataset with VGG16, VGG19, ResNet50, and MobileNet top four best performing models. In Recall scores, Inception V3 bagged 100%, followed by DenseNet, VGG16, and Xception, on the TI-CNN dataset. For the EMERGENT dataset, in terms of accuracy scores, ResNet50 and Xception secured 51.26% each (highest accuracy), followed by DenseNet and MobileNet with 48.65% 46.93%, respectively. VGG16 performed better on TI-CNN, whereas ResNet50 and Xception on the EMERGENT dataset indicate varying importance and reliance on different Image-CNN models regarding variations in the dataset. We show the performance of the eight Image-CNN models on the Image-only dataset MICC-F220. Xception with 100% accuracy, followed by VGG16 with 95.05% accuracy, VGG19 with 91.97%, and AlexNet with 91.54% accuracy, lead the table.

We provide the final output performance figures of the proposed Coupled ConvNet framework on the two datasets. Comparisons based on Accuracy, Precision, Recall, and F1 scores can be inferred from the table. To eliminate complexity in deciphering the best model or the most relevant text and Image multi-modal fake news detection, let us analyze the Accuracy score comparisons between the TI-CNN and EMERGENT datasets. The combination of Text-CNN with VGG16 performed the best on each of these datasets with 98.93% and 94.05% scores, respectively. While Text-CNN and VGG19 combination performed with 98.4% accuracy on TI-CNN as the second best, Text-CNN and MobileNet coupled ConvNet produced 93.98% accuracy on the EMERGENT dataset, being the second-best. Third and fourth-best performance on TI-CNN was observed with DenseNet and InceptionV3 with 97.86% and 97.65% accuracy, respectively, and on EMERGENT, ResNet50, and Xception produced 91.47% and 90.98% accuracy, respectively.

Weights produced the best classification results can be concluded to be 0.5 for both text and image. Text and image both offer an equal contribution to detecting fake news efficiently. In some cases, the participation can be discovered to be 7:3 for text and image data modalities. It highlights text being a necessary component for fake news detection. It is also evident that exploring visual modality is equally essential.

The MICC-F220 dataset consists of tampered and unaltered images. Images under the unaltered category have not been edited in any form, and thus it serves the purpose of efficiently

distinguishing between real and fake images. We deduce that CNN models are highly accurate in detecting fake news where the text is classified based on their vector embeddings and images have been tampered with or edited. We propose using combinations of text and image CNN models to detect fake news using multiple textual and visual modalities. Hence, we provide performance comparisons of these models to make a witty selection for counterfeit news detection tasks. The accuracy obtained with the MICC-F220 dataset is as high as 100% using XceptionNet, and the lowest is 59.52% with the ResNet50 model. Other models have also demonstrated outstanding performance with high accuracy values. This performance highlights the need for larger visual and multi-modal datasets with distinguishable latent features.

It can be concluded that VGG16 is a consistent performer. Xception and MobileNet are observed to be the next best performers. Despite achieving 100% result with the MICC-F220 dataset, Xception displays average performance with the other two datasets. It can be regarded as being slightly inconsistent with datasets.

Table 8: Performance of Text-CNN Module on TI-CNN and EMERGENT

Dataset	Accuracy	Precision	Recall	F1-Score
TI-CNN	96.26	95.77	96.00	95.89
EMERGENT	93.56	94.07	89.35	93.12

Table 9: Performance of Image-CNN Module on TI-CNN and EMERGENT

Image Model	TI-CNN				EMERGENT			
	Accuracy	Precision	Recall	F1Score	Accuracy	Precision	Recall	F1Score
ResNet50	77.54	58.22	88.57	70.25	**51.26**	**58.59**	**76.82**	**66.26**
VGG16	82.72	63.49	97.65	77.26	45.18	51.29	58.56	54.77
VGG19	81.04	59.77	88.98	71.32	41.90	48.42	42.45	45.00
InceptionV3	58.76	09.18	100.00	16.81	43.54	50.28	54.41	52.89
DenseNet	60.00	11.40	97.96	20.43	48.65	52.93	63.71	57.25
Xception	62.57	10.51	97.62	18.98	**51.26**	57.19	68.42	62.59
AlexNet	59.44	48.32	91.69	59.87	43.62	50.71	48.64	49.71
MobileNet	73.37	55.66	79.46	65.46	46.93	55.18	52.53	53.48

Table 10: Performance of Image-CNN Module on MICC-F220

Image Model	MICC-F220			
	Accuracy	Precision	Recall	F1-Score
ResNet50	61.36	61.37	59.52	60.43
VGG16	95.05	95.15	78.26	85.88
VGG19	91.97	92.02	83.33	87.46
InceptionV3	91.01	91.01	90.63	90.82
DenseNet	89.63	89.61	92.00	90.79
Xception	100.00	100.00	93.75	96.78
AlexNet	91.54	91.52	95.00	93.22
MobileNet	82.82	82.73	100.0	90.55

Table 11: Performance of Coupled ConvNet Model on TI-CNN and EMERGENT

Text Model	Image Model	TI-CNN				EMERGENT			
		Accuracy	Precision	Recall	F1Score	Accuracy	Precision	Recall	F1Score
Text-CNN	ResNet50	96.90	96.48	96.71	96.59	91.47	91.90	88.95	87.96
	VGG16	98.93	98.21	99.22	98.71	94.05	90.08	86.72	86.12
	VGG19	98.40	97.18	98.96	98.06	89.12	89.11	85.80	85.49
	InceptionV3	97.65	97.65	97.19	97.42	89.88	89.63	86.44	85.88
	DenseNet	97.86	98.34	96.96	97.64	90.64	90.35	87.39	86.73
	Xception	97.22	94.36	98.92	96.59	90.98	91.77	88.02	87.97
	AlexNet	96.91	96.26	96.94	96.60	89.66	89.80	86.09	86.02
	MobileNet	97.54	97.41	97.18	97.29	93.98	91.48	86.22	86.55

Table 12: Accuracy Comparison of Image-CNN models on all datasets

Image Model	TI-CNN	EMERGENT	MICC-F220
ResNet50	77.54	51.26	61.36
VGG16	82.72	45.18	95.05
VGG19	81.04	41.90	91.97
InceptionV3	58.76	43.54	91.01
DenseNet	60.00	48.65	89.63
Xception	62.57	51.26	100.00
AlexNet	59.44	43.62	91.54
MobileNet	73.37	46.93	82.82

We conclude that CNNs perform better when the dataset is comprised of all tampered images. Data with fake images where fake corresponds to false, tampered, old, misleading, and unrelated images perform somewhat lower as CNNs could detect only latent features. For utilizing features contained in all types of fake photos, multi-modal frameworks are needed which can incorporate elements contained in all kinds of counterfeit images. The above best performing models are likely to show better performance over larger training datasets.

Baseline Comparison

We validate our results with both single modality textual and visual methods and multi-modal methods for a fair comparison of our proposed work with established baselines. We compare the results for each dataset separately. The proposed task being the first to examine Emergent on a visual basis, we establish a baseline for visual and multi-modal fake news detection on this dataset. Due to the absence of work performed in the visual domain, this task stands first, and hence, the comparison is provided for textual classification. The results for comparison are noted from those mentioned in the existing literature.

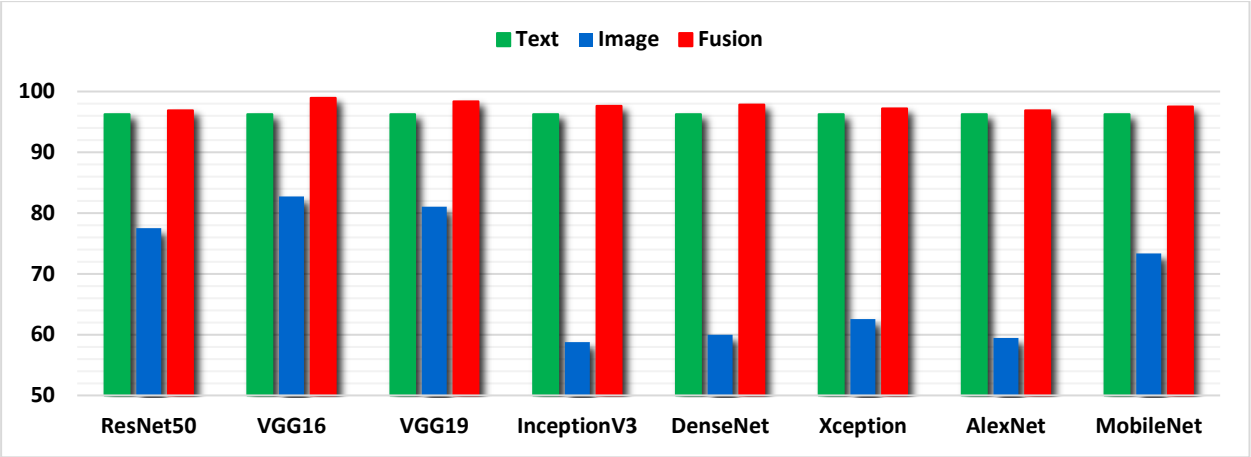


Figure 27: Accuracy Comparison on TI-CNN dataset

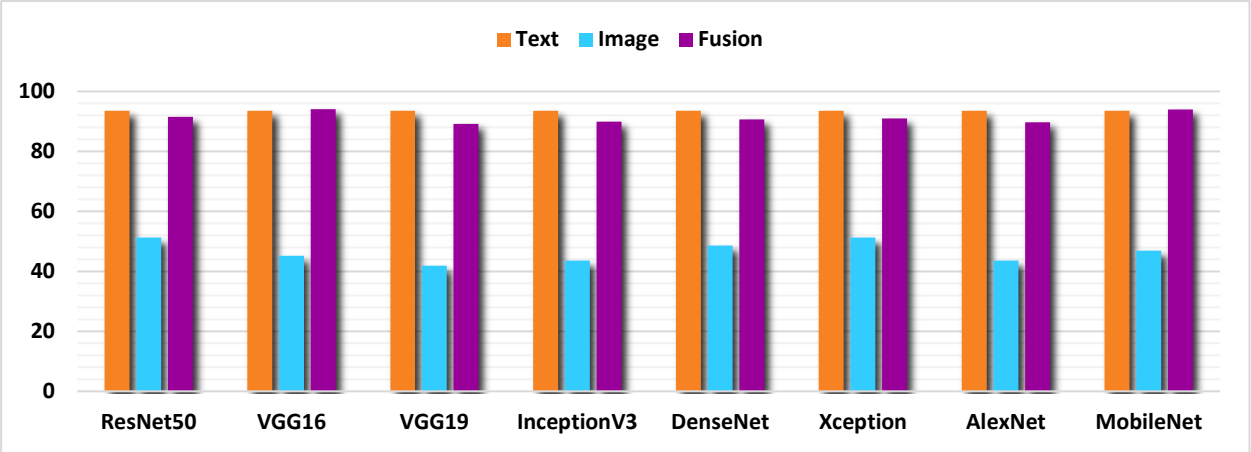


Figure 28: Accuracy Comparison on Emergent Dataset

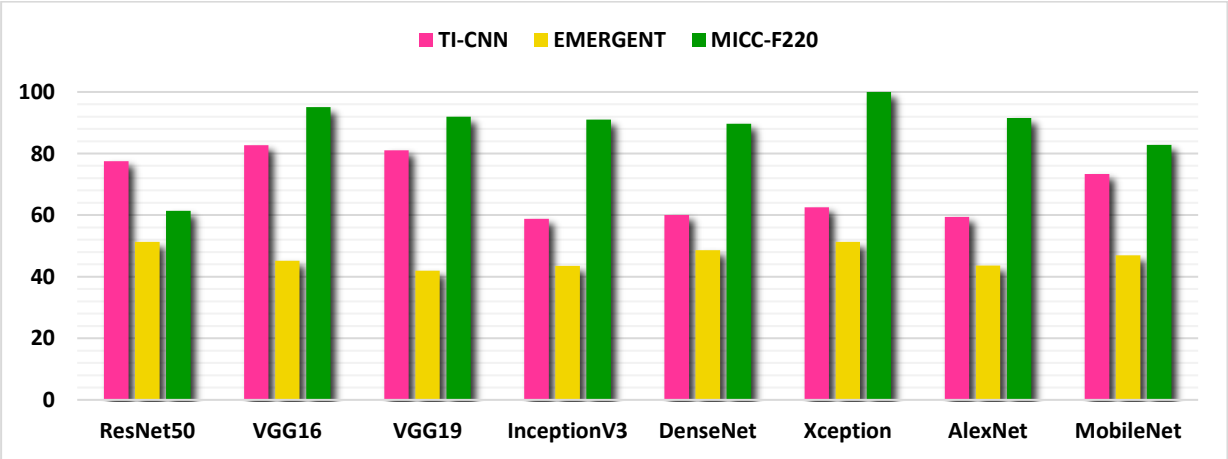


Figure 29: Accuracy Comparison of Image-CNNs on three datasets

Table 13: Baseline comparison of TI-CNN dataset

Modality	Baseline	Method	P	R	F1
Textual	(Yang et al.) [1]	LR	57.03	41.14	47.80
		GRU	88.75	86.43	87.58
		LSTM	91.46	87.04	89.20
		Text-CNN	87.22	90.79	88.97
	Proposed Method	Text-CNN	**95.77**	**96.00**	**95.89**
Visual	(Yang et al.) [1]	CNN-image	53.87	42.15	47.29
	Proposed Method	ResNet50	58.22	88.57	70.25
		VGG16	63.49	97.65	**77.26**
		VGG19	**59.77**	88.98	71.32
		InceptionV3	09.18	**100.00**	16.81
		DenseNet	11.40	97.96	20.43
		Xception	10.51	97.62	18.98
		AlexNet	48.32	91.69	59.87
		MobileNet	55.66	79.46	65.46
Textual and Visual (Combined)	(Yang et al.) [1]	TI-CNN	92.20	92.77	92.10
	Proposed Method	ResNet50	96.48	96.71	96.59
		VGG16	**98.21**	**99.22**	**98.71**
		VGG19	97.18	98.96	98.06
		InceptionV3	97.65	97.19	97.42
		DenseNet	98.34	96.96	97.64
		Xception	94.36	98.92	96.59
		AlexNet	96.26	96.94	96.60
		MobileNet	97.41	97.18	97.29

TI-CNN: On this dataset, Yang et al. experimented with multiple text classification methods: Logistic Regression, GRU, LSTM, and Text-CNN [1]. For the visual domain, Yang et al. used image-CNN with a proposed architecture of convolutional layers. They created a TI-CNN dataset and performed text classification using the embedding layer and one-dimensional convolutional layer. Image convolution is achieved by using a model that contains three convolutional layers. Filter size is kept as $\mathbf{3 \times 3}$. Thirty-two filters have been used, and the layers inculcate the ReLU activation function. All of our text and image models surpass the scores obtained by Yang et al. [1]. Individual text and image models proposed by us provide accuracies higher than those observed by Yang et al. In the multi-modal aspect, our approach obtains the highest F1-score of 98.71% using a combination of Text-CNN and VGG-16, which outperforms the state-of-the-art result by ~6%. It establishes the proposed work as a new baseline for multi-modal fake news detection.

Table 14: Baseline comparison on EMERGENT (FNC) dataset

Modality	Baseline	Method	Acc%
Textual	(Conforti et al.) [37]	Bi-LSTM	33.00
	(Bourgonje et al.) [38]	LR	89.59
	(Thorne et al.) [39]	Ensemble Method	90.89
	Our Method	Text-CNN	**93.56**

Table 15: Baseline comparison of MICC-F220 dataset

Modality	Baseline	Method	TPR%	FPR%
	(Uliyan et al.) [41]	Hessian Method	92.00	08.00
	(Uliyan et al.) [42]	Blur Detection	96.50	02.86
	(Doegar et al.) [40]	AlexNet	100.0	12.12
	(Amerini et al.) [36]	SIFT	100.0	08.00
	Our Method	ResNet50	59.52	**0.00**
		DenseNet	78.26	**0.00**
		AlexNet	83.33	**0.00**
		InceptionV3	90.63	83.33
		VGG16	92.00	**0.00**
		MobileNet	93.75	25.00
		VGG19	95.00	12.50
		Xception	**100.0**	**0.00**

EMERGENT: Experiments previously performed by researchers used FNC (FakeNewsChallenge) dataset, which has been derived from Emergent. We compare text-classification results of our model with the LSTM model used by Conforti et al. [37], Logistic Regression applied by Bourgonie et al. [38], and an ensemble of multiple methods deployed by Thorne et al. [39]. Usage of the Text-CNN classification model beats these established baselines, providing an accuracy of 93.56%. Visual fake news detection on this dataset has not been performed previously as the dataset was limited to textual information only. We leverage the task to a visual analysis by adding images extracted from page websites and provide a maximum of 51.26% accuracy using ResNet50 and Xception models.

MICC-F220: Earlier tasks on this dataset have incorporated image forgery detection techniques with Amerini et al. [36] demonstrating 100% TPR and 8% FPR. Most of our proposed model methods have displayed 0% False Positive Rate, and XceptionNet provides 100% True Positive Rate outperforming all other baselines. 0% FPR demonstrates that no fake samples were wrongly classified as real during the testing phase, and 100% TPR shows that all unaltered samples in the test set were classified into the correct class. A model that achieves 0% FPR and 100% TPR is a perfect classifier. With the proposed approach, the Xception model is the ideal classifier for this dataset, classifying all test samples into correct classes.

3.4 SUMMARY

A novel Coupled ConvNet architecture is proposed comprising of Text-CNN and Image-CNN modules. This work accomplishes fake news detection using several convolutional

models on text and image data. Our first contribution provides image datasets for counterfeit news detection, which we have publicly available on Kaggle. We compare the performances of image classification models, namely AlexNet, ResNet50, DenseNet, MobileNet, Xception, InceptionV3, VGG-16, and VGG-19, on three real-world datasets TI-CNN, EMERGENT, and MICC-F220. Text-CNN module has been used over TI-CNN and EMERGENT and Image-CNN module on all of the above datasets. We have trained these models and obtained their training, validation, and testing accuracy scores. We utilized latent features for fake image classification and analyzed how well classification can be performed, comparing various efficiencies. All of our models have surpassed fake news detection baselines with high results. The proposed architecture provides a new fake news detection method using convolutional neural networks and establishes a new baseline in this domain. The source codes of the proposed work have been made publicly available. Our proposed model would function more efficiently on larger datasets. We intend to apply these models to larger datasets further. We are also motivated to tune further the parameters used in these models to enhance classification accuracy. Additionally, we focus on coming up with an efficient classification model based on CNN's with fine-tuned hyperparameters serving greater accuracies and better fake news detection.

CHAPTER 4

ARCNN FRAMEWORK

4.1 INTRODUCTION

COVID-19 is a fatal pandemic and has breathtakingly raised the infodemic associated with it. 'Infodemic' is a term coined by the World Health Organization to describe the spread of false news in enormous amounts at the time of coronavirus pandemic. Verbal rumors spreading through individuals since the origination of humankind have camouflaged into online fake news from person to mass, all credits to feasible technological access. After the 2016 US presidential elections, the pandemic has appeared as one of the most significant events of misinformation propagation where each individual on the internet has been the source or the consumer of misinformation [122]. Social media and networks have provided a platform for netizens to let the infodemic move incessantly. News gets manipulated several times when it travels through word of mouth [123]. On online platforms, such misinformation leaves behind traces in the form of big data. Fake news is not a technical issue in the media but rather a deliberate human activity [124]. Technology has rendered us to detect the authenticity of data on the internet. Information spread on the internet in the form of multimedia became progressively increasing with the intent of reaching out to a larger audience. Visuals bypass human minds more promptly than long and often dull texts and leave a lasting impact [125]. Users on social media have varied ideologies. Each user perceives information differently depending on several factors like education, personal background, political stand, religious inclination, and demographics [126]. The information thus gets manipulated several times in the course of reaching the people [127]. Social Media users with malicious intent are using multimedia as a tool to spread false information. With the aim of technological advancement to nurture human lives, such betterment also pulls us back in the form of challenging regressions. This builds up challenges in the field of fake news detection [128]. The fake news detection task has become more complex because detecting visual information is more complicated than plain text. Such detection tasks are being performed with the help of deep learning techniques [129].

False news can be broadly divided as misinformation (referring to news that people spread unaware of its credibility) and disinformation (false news mainly spread with a defined motive) [130]. In the current unpredictable global scenario, people have become desperate to

grasp as much information as possible. Such misinformation causes misconceptions in people's minds and causes severe impacts [131]. Being a topic of concern, many have stepped forward to impart as much as they get to know, either being informed or ill-informed. Credibility analysis and fact-checking of every single piece of information on the internet is not feasible. In such a peculiar time, when it is of utmost importance to deliver authentic information to mass, the internet is flooded with false news that links coronavirus to a wide range of entities [132].

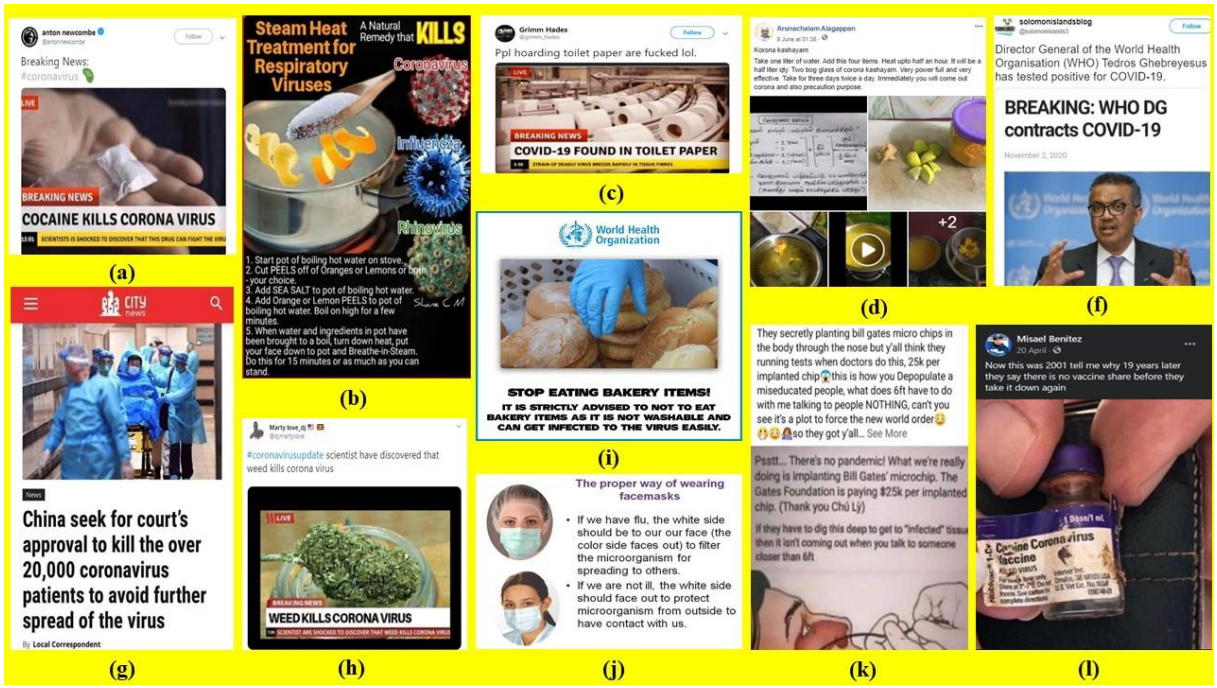


Figure 30: Examples of fake news related to COVID-19

Infodemic transmitted as rapidly as Covid-19 itself, in some cases, faster, owing to the advanced internet technology and online social platforms [133]. It became a raging issue emanating from conspiracy theories, political agendas, fake advisories, and more. Several false claims stating multiple remedies as a cure of the disease evolved, misleading people into self-medication and unproven treatment procedures [134]. Figure 21 is a representation of a few of the examples of fake news related to coronavirus. These screengrabs have been collected from social media platforms, and the news has been verified and declared false or misleading by official fact-checking sites. These examples show fake remedies suggested by people to cure coronavirus. Figure 21 (i) shows a fake claim attributed to the World Health Organization that advises people to stop eating bakery items. Various fake claims circulated on the internet state that the advisory or prevention mechanism has been issued by the WHO or various other

reputed official organizations. Misleading information has gripped the multimedia scenario claiming the cure for the disease textually as well as visually. There is high urgency to provide a curb to such malefic instances. Visuals catch one's attention more promptly and are easier to comprehend, unlike text which requires conceptual understanding. This makes the study of multimodal content critical. We, therefore, design a framework that exploits both textual and visual matter to perform the classification of fake and real news.

Overwhelmed by the enormous amount of fake news pouring amidst the pandemic, we were encouraged to design an architecture that discerns misleading information based on the features inherent within them. The main goal of this work is to establish a unified framework that alleviates fake news detection tasks to help mitigate the infodemic. We propose the ARCNN model (Allied Recurrent and Convolutional Neural Network) to distinguish COVID-19 related fake and real news. Studies relying on multimodal information for online news verification are limited. Existing research is primarily focused on text-based fake news detection utilizing traditional machine learning algorithms. The primary drawback observed in machine learning algorithms is that they require a manual feature extraction process. The proposed ARCNN overcomes this limitation by incorporating deep learning architectures that automatically learn feature extraction using neurons during the model training phase. Another gap observed in machine learning algorithms highlights their inability to mine inherent features within the information. In contrast, our framework holds the advantage of recognizing patterns in the data provided to them, such as identifying the writing style in text or recognizing image tampering in the images. Also, to handle large volumes of data, deep learning-based ARCNN is more effective where machine learning-based frameworks fail to perform.

Visual data is an essential contributing factor in the spread and detection of fake news. Unlike the detection works performed until now in the infodemic detection domain, our framework uses multi-modal features from COVID-19 related discussions on the web and fuses them to generate classification predictions. Inspired by the previous multi-modal approaches [135, 136, 137] we utilize RNN and CNN architectures and fine-tune them to fit on coronavirus related texts and images precisely. Our architecture relies upon inherent textual and visual features that the ARCNN model efficiently exploits. It focuses on knowledge-based detection in the text domain, which analyses the writing style differences of fake and real news, and secondly, self-extraction of visual features by CNNs for efficient image classification. The choice of using improved RNN models, namely LSTM and Bi-LSTM, is to mine the advantage of their high data retaining the capacity for sequential inputs. The proposed RNN sequences in

this work are capable of extracting useful long-term dependencies in textual data. The proposed LSTM and Bi-LSTM networks can learn textual patterns observed in fake and real news on a sentence level. The lags between the occurrences of similar patterns are remarkably handled and used for classification by LSTMs. Also, the proposed networks overcome the vanishing gradient problem commonly encountered in traditional RNNs. The RNN stream of the proposed ARCNN serves to detect valuable patterns in fake news by identifying its writing style. The usage of CNNs for image classification is supported by the fact that they can identify inherent features within an image. In addition to being computationally efficient, CNNs can recognize distinctive features existing in images of different classes. The CNN architectures and optimization in the proposed ARCNN offer high adaptability to any input data. In order to build a multimodal approach that uses both textual and visual information present in an online post, the RNN and CNN models are needed to be combined. Researchers perform such a combination using two primary techniques: early fusion and late fusion [138]. As suggested by their names, the combination is performed at an early stage prior to training the deep learning model in case of early fusion, whereas, in late fusion, the features extracted are combined after training each of the models separately. Early fusion is performed by concatenating the features obtained from each model. To perform late fusion, we employ four techniques: sum, max, average and weighted average. Their respective mathematical operations support the fusion of features from different models. Early fusion is a complex operation, whereas late fusion is relatively easier to perform [138]. However, the usage of early fusion results in lower computation time as training is performed only once. Late fusion, being relatively more straightforward, takes longer training durations. To explore the effects of both fusion mechanisms, we develop two variants of the ARCNN framework. Thus, combining the proposed RNN and CNN pipelines, we present one of the earliest multi-modal frameworks for infodemic detection. A summarization of the contributions of this work are provided as follows:

1. We introduce CovID, a multi-modal Coronavirus Infodemic Dataset consisting of over 3500 real and fake news with text and images.

2. We propose novel ARCNN architecture that incorporates proposed RNN and CNN models. The RNN stream is experimented with by the use of LSTM and Bi-LSTM. To experiment with various CNN architectures, we propose a novel CNN model. We also use four pre-trained CNN models, VGG-16, InceptionV3, Xception, and MobileNetV2, fine-tuning them to achieve high performance for fake news classification.

3. We experiment with five methods to fuse text and image modalities using early fusion and late fusion. The early fusion mechanism in our approach performs simple concatenation of features. The late fusion variant uses average fusion, weighted average fusion, sum fusion, and max fusion techniques for combining the RNN and CNN models.

4. The performance of the proposed ARCNN model is evaluated by experimenting with multiple combinations of RNN and CNN models on six multi-modal COVID-19 fake news datasets that include ReCOVery, CoAID, MediaEval 2020, and our proposed CovID.

5. Our work analyses the percentage contribution of textual and visual features in misinformation detection.

6. Evaluation results are presented in terms of a wide range of metrics to present an exhaustive analysis based on accuracy, precision, recall, F1-score, roc-auc score, FPR, specificity, and MCC.

The outcomes of this research are:

1. Bi-LSTM is observed to be a better RNN choice over LSTM for textual fake news detection as it performs marginally better. In visual classification, XceptionNet is the leader with the highest maximum, average, and minimum accuracy, followed by MobileNetV2, VGG-16, InceptionV3, and the proposed CNN model.

2. It is observed that weighted average fusion results in the highest accuracies, followed by early fusion, average fusion, sum fusion, and max fusion. This explains that the framework works best when text and images are assigned a suitable weight while combining the modalities. Though being complex, early fusion stands to be the second-best combinatorial method with a reduced runtime.

3. It is observed that visuals play a critical role in fake news identification. The weighted average fusion demonstrates a 30-50% contribution of images towards infodemic detection.

4. Through the experimentations performed on distinctive datasets that contain posts from news articles and social media, it is observed that the corpora constitution plays a vital role in infodemic detection, signifying the influence of writing style on detection mechanisms. We observe that the proposed framework performs better classification on social media posts than on complex and fairly long-written news articles.

4.2 PROPOSED DATASET

The infodemic rose at an alarming rate as the pandemic spread its wings over the globe, and given the extreme worries to curb the disease, fighting with an infodemic while global chaos is in existence has become quite challenging. There is a scarcity of multi-modal infodemic datasets, which is crucial for developing fake news detection systems. Researchers worldwide have responded impulsively to understand the complexities and have acted promptly to introduce various infodemic datasets and detection methodologies. Shahi and Nandini have introduced one such repository, FakeCovid, a multilingual collection of fact-checked news across 105 countries [139]. Their dataset motivated us to create CovID, our multi-modal dataset for textual and visual fake news detection. Rather than using FakeCovid to extract visual features, we decided upon extracting the news articles from scratch. This aided us in building a dataset of a more extensive date range from 04/01/2020 to 30/10/2020. Another limitation we encountered in FakeCovid was the biased nature of the dataset with very few numbers of authentic or reliable articles. To overcome the limitations, we proposed to build CovID from scratch extracting real and fake news items along with their visual content. We have extracted data from various news sources like news websites, fact-checking websites, and Twitter. To create a balanced dataset, we used the following sources for each label:

Poynter: Poynter Institute maintains the International Fact-Checking Network in the view of debunking false news across the world. During the infodemic, they maintain a database of coronavirus-related fact-checked news articles in more than 40 languages from websites of several countries in the world. We started by scraping page URLs of fact-checked articles listed on the Poynter website. Beautiful Soup aided the extraction, a python library used to crawl elements from web pages. After getting the URLs of all fact-checked articles, we use them to crawl various details in the dataset, most important of which are news title, news text, and image URL. We merged various strong and weak categories of false news under the fake label. These categories are: false, false context, conspiracy, false headline, inaccurate, incorrect, mainly false, misleading, primarily false, pants on fire, partially false, and partly false. The false information debunked in the fact-checking articles is a mix of social media posts, contributed mainly by Facebook and Twitter users, and malicious websites are posting false claims. This set builds up our data under the False label category with news titles, text, and image links.

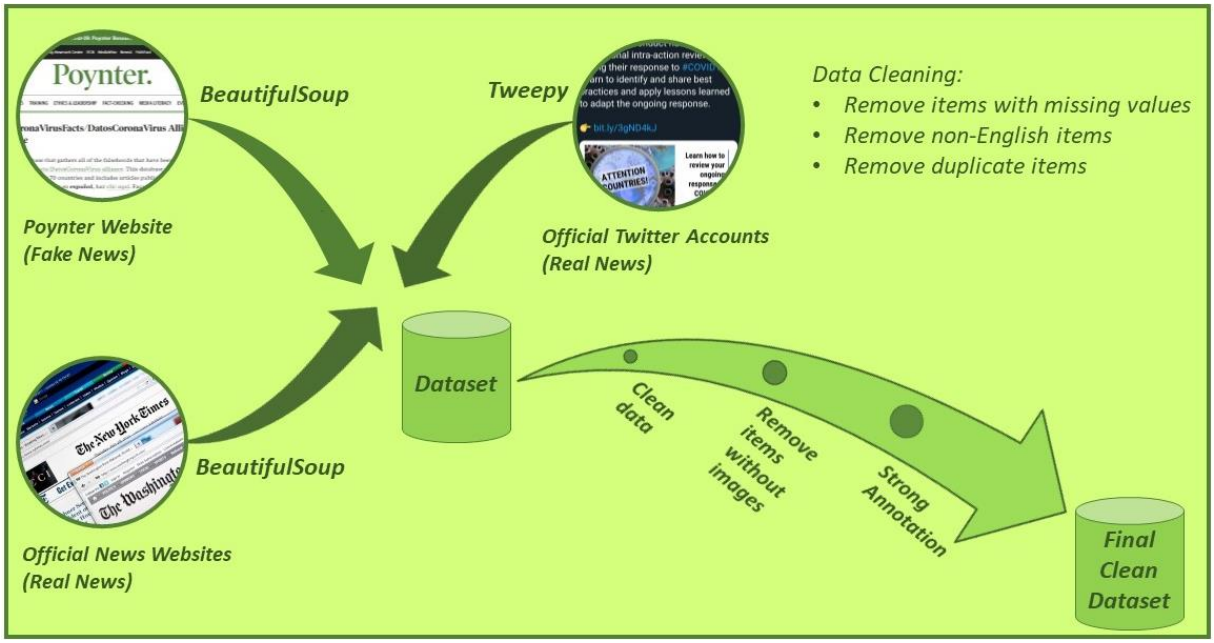


Figure 31: Data collection and pre-processing workflow

Official News Websites: Deep learning algorithms learn on training data to be able to distinguish between classes. This generates the need for well-classified data under different labels. We received a meager count of true articles from the Poynter website, due to which we shifted to collecting true articles from official sources of news. We created a list of official news websites that are providing trustworthy news in the times of COVID-19. The extraction process is the same as extracting false news articles. We used each website and collected news articles that were linked with coronavirus. The keywords used were COVID-19, COVID, and coronavirus. We obtained a collection of true news titles, text, and image URLs.

By examining the collection thus obtained, we figured that data under the false label is a mix of false news articles from websites and social media posts. In contrast, the data under the true category contains only official news articles. Knowledge-based detection, which we propose to apply in our framework, works upon the writing style of the text. It focuses on how sentences are structured and words are linked together. Paying heed to this minute detail of the technique, we perceive that official news is structured formally and in a well-defined way rather than social media posts with inconsistent writing styles. To provide a bias-free detection, we decide to contrast our collected false news with an unbiased mix of news articles and social media posts. To proceed in this direction, we extract social media's true news from Twitter. These contrasting datasets assist us in inspecting the effect of corpora in fake news detection tasks. We call these two versions of our dataset CovID I and CovID II.

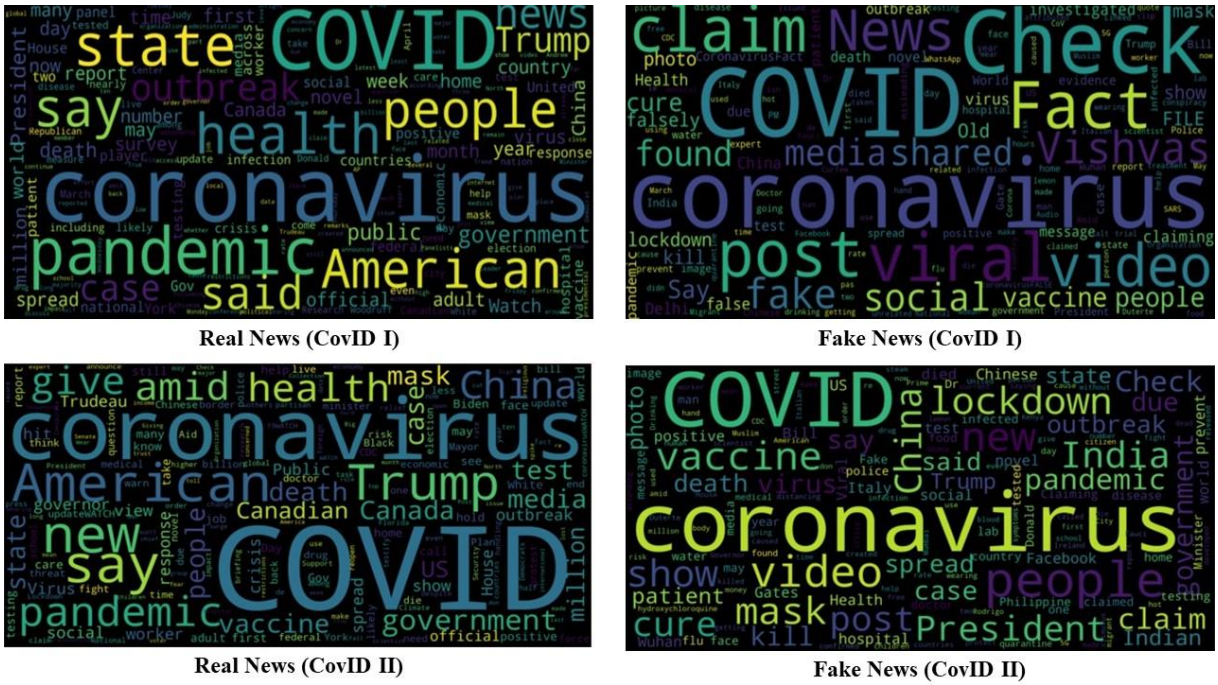


Figure 32: Wordclouds of real and fake news from CovID I and CovID II

Twitter: The extraction process is supported by Twitter REST API using Tweepy to extract historical tweets. We shortlisted such official Twitter users who provided authentic information during the times of Covid and fetched all of their tweets since dated back to January 01, 2020, till October 30, 2020. The extraction process provided us with multiple information, of which we have used tweet texts and image URLs.

Pre-processing: The pre-processing steps involve many data cleaning steps to filter out unwanted data. At first, we removed all the multilingual data, keeping our repository with solely English news. We then removed all news items that did not contain visual information. For multi-modal detection, we kept only news articles with images along. We then removed duplicate items and any rows containing missing values. After performing a content analysis of the remaining data, we strongly labeled our dataset by manual annotation. The items were weakly labeled as true or false depending on their extraction sources. For final confirmation of their classification, two annotators went through each item in the dataset and provided a strong label based on the mutual decision to the items. The annotation is supported by inter-coder reliability where the annotators verify the labels by authenticating the headlines, text and images. To stay updated to the continuously evolving scientific knowledge about the COVID-19 facts, the datasets are regularly being fact-checked and verified. Word clouds for real and fake classes of both the proposed datasets are shown in figure 23.

4.3 PROPOSED METHODOLOGY

We have envisioned the fake news detection task as a combination of text and image classification. Deep learning is widely being used and is proving effective in such tasks. We propose ARCNN, Allied Recurrent and Convolutional Neural Network, which uses an RNN model for text classification and CNN for image classification. We introduce two variants of the ARCNN model, which differ on how the text and image modalities are combined for compelling predictions. The workflow is illustrated in figure 25. The architecture diagram for ARCNN is presented in figure 24, depicting early fusion and late fusion variants of the proposed ARCNN architecture.

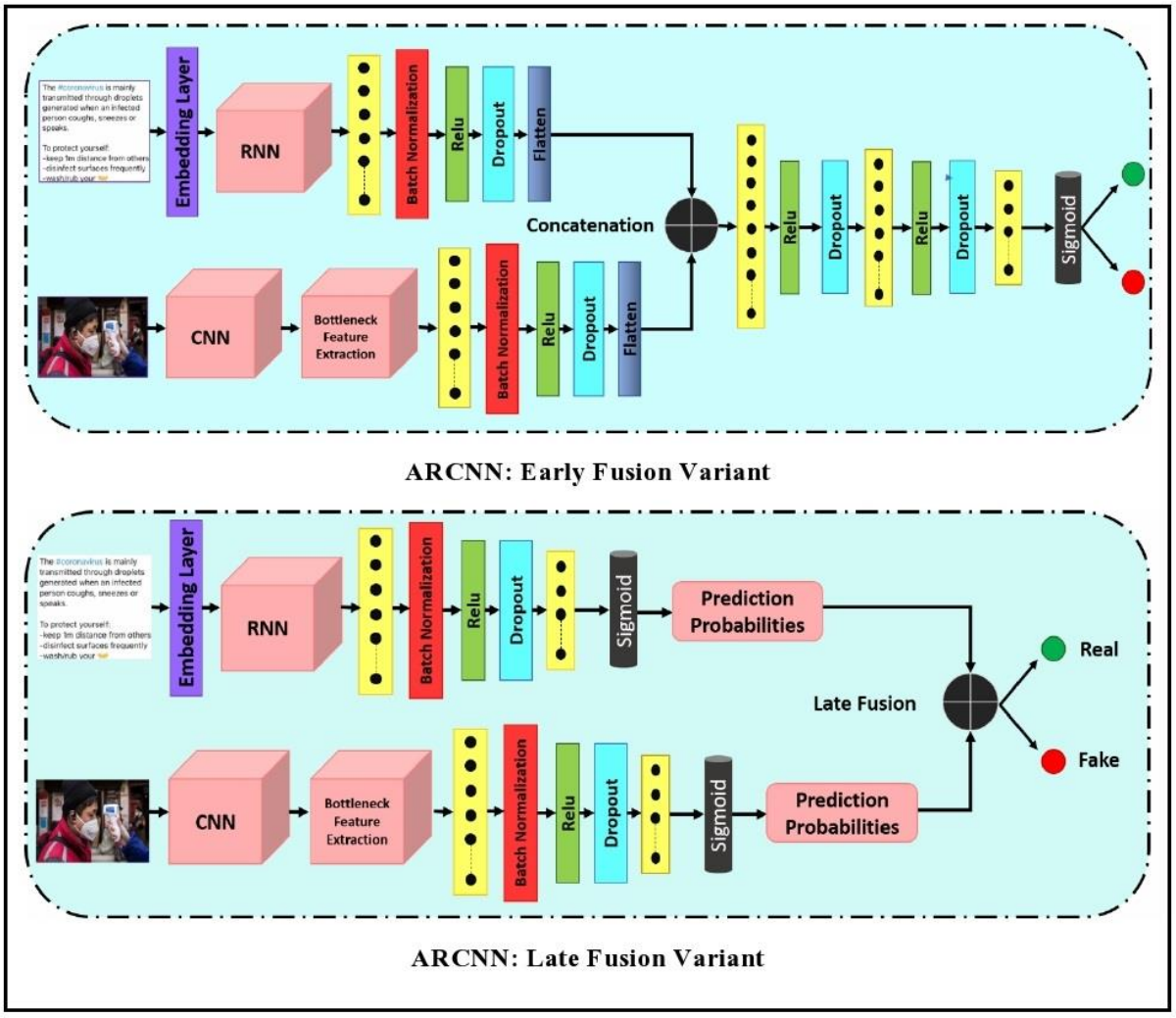


Figure 33: ARCNN Architecture Diagram

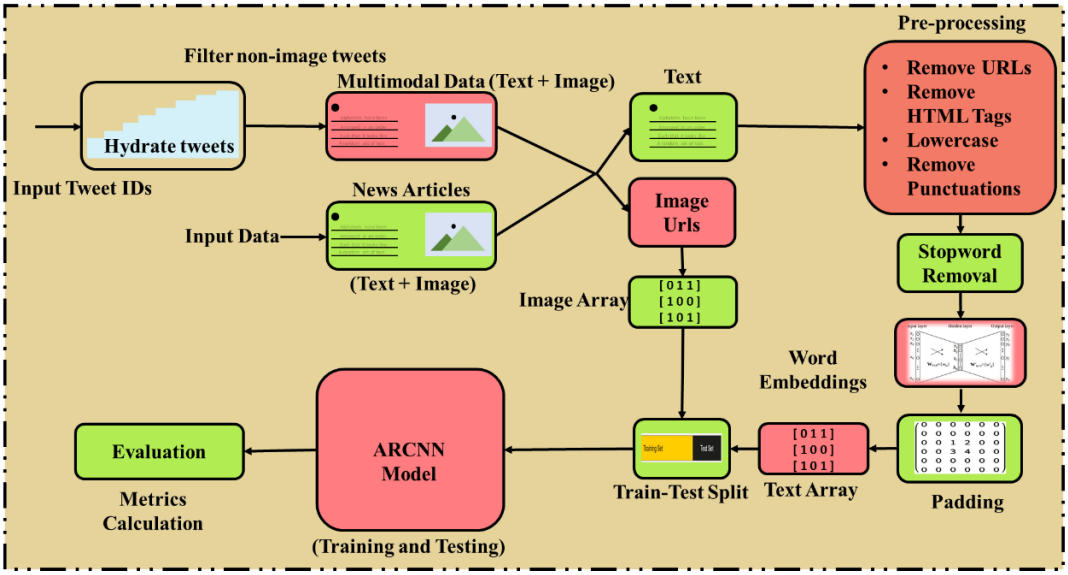


Figure 34: Workflow of the proposed methodology

RNN Component

The selection of RNNs for text classification is based on the advantage of having a memory base. Unlike simple neural networks, the input of the current layer intakes the output of the previous layer forming a connection that remembers previous sequences and helps predict the next step. All the hidden layers in an RNN can be merged as one recurrent layer. RNNs have proved to be extremely important as they can process information of arbitrary length and remember this information throughout the recurrent states. Despite such capability of traditional RNNs, the vanishing gradient problem is encountered in their use. Such an issue arises because the RNN allocates a deeper memory for recent input signals than the previous ones [140]. The problem is resolved using backpropagation through a particular type of RNN known as Long-Short-Term Memory Networks (LSTMs). A representation of an LSTM network is given in figure 26. The top horizontal line in the figure is known as the “cell state” responsible for storing and removing information. LSTM networks incorporate a gate mechanism by using an input gate (i_t), output gate (o_t), and a forget gate (f_t). The gated structure of this new RNN overcomes the problem of traditional RNNs. These gates perform pointwise multiplication to process the input information.

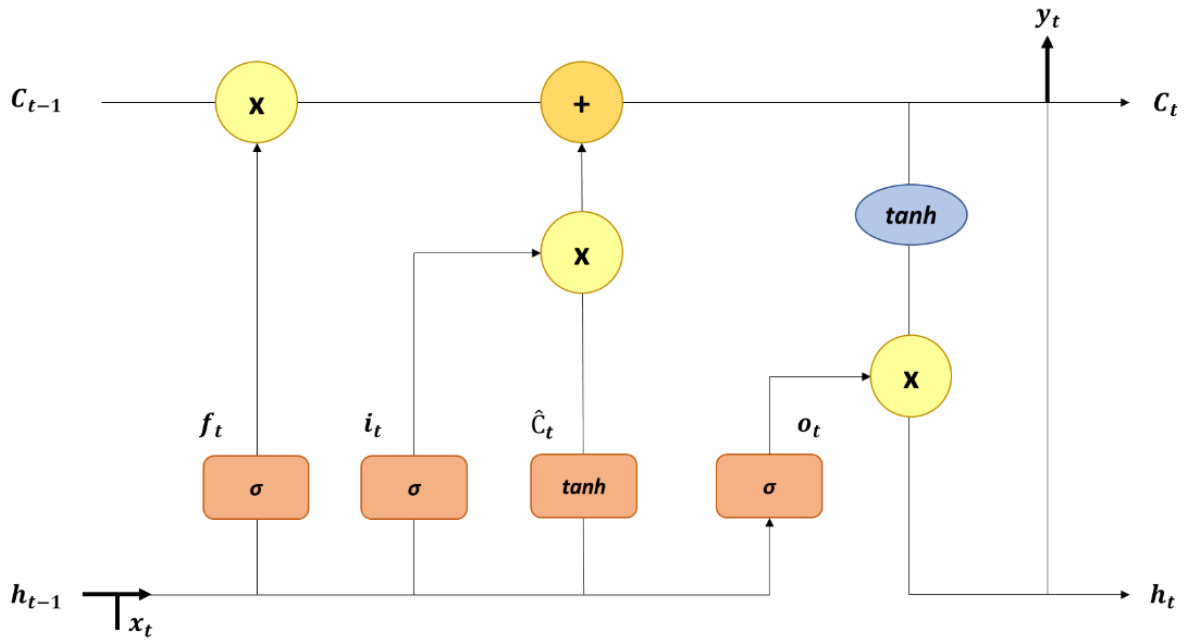


Figure 35: Architectural representation of LSTM

For inputs given as $x_{t-1}, x_t, x_{t+1}, \ldots, x_n$, the current state in an LSTM is calculated as $h_t = f(h_{t-1}, x_t)$, where h_{t-1} is the previous state and h_t is the current state. The forget gate (f_t) selectively chooses which information must be transferred to the following cell states. It is mathematically represented as:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

where σ denotes the activation function. W_f and b_f represent the weight and bias at a given time *t* at the forget gate layer. Next, the input gate is responsible for deciding which information is to be stored in the cell state given by Eq. 2

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

where, W_i and b_i are the weight and bias for the input gate.

The activation function, the sigmoid function produces a vector $\hat{C}_t$, defined by Eq. 3

$$\hat{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

The previous cell C_{t-1} is updated to C_t using Eq. 4

$$C_t = f_t * C_{t-1} + C_{t-1} + i_t * \hat{C}_t \quad (4)$$

The final cell state is responsible for providing the output o_t of the network, which is defined by:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

where W_o and b_o are the weight and bias at the output layer.

A bidirectional LSTM or Bi-LSTM network adds to the advantage of a simple LSTM network. While LSTM is unidirectional and can store only past information in its cell states, a Bi-LSTM directs information forward and backward. Its architecture is illustrated in figure 27. For a given sequence of inputs $x_{t-1}, x_t, x_{t+1}, \ldots, x_n$, the output from the forward layer $\vec{h}$I s calculated, whereas for a reverse sequence, $x_n, x_{n-1}, x_{n-2}, \ldots, x_{t-1}$, the output $\overleftarrow{h}$ is calculated through the backward layer using Eq. The output of the Bi-LSTM network is denoted as:

$$Y_T = y_{t-1}, y_t, \ldots, y_{t+n} \tag{6}$$

where, $y_t = \sigma(\vec{h}, \overleftarrow{h})$ and σ is a concatenation operation.

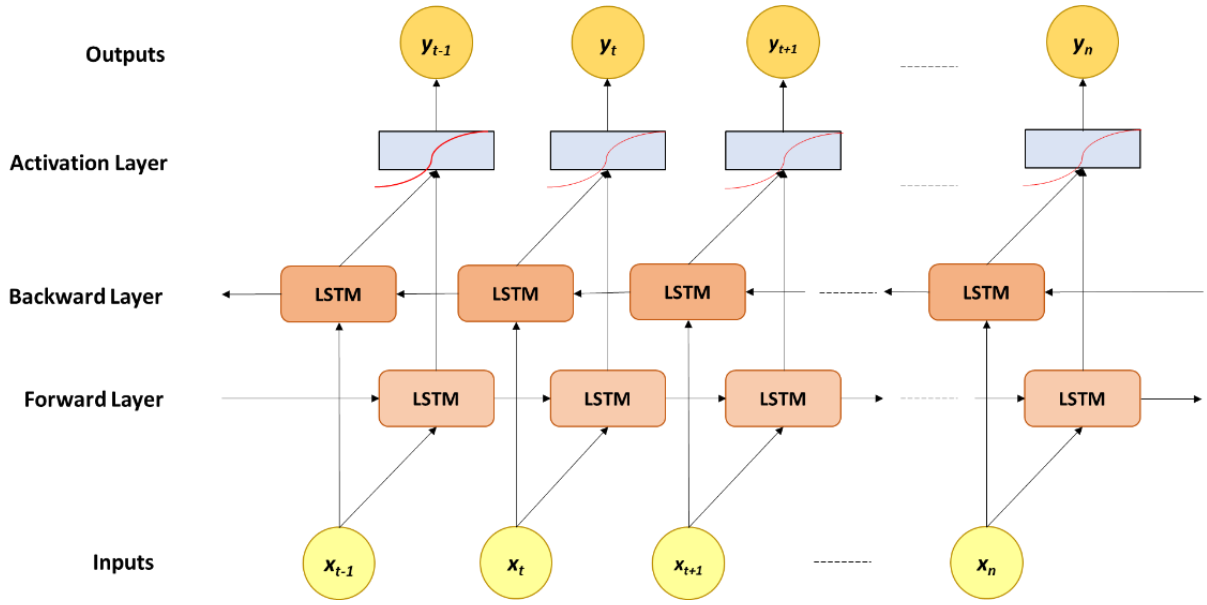


Figure 36: Architectural representation of bidirectional LSTM

Text input provided to the proposed ARCNN goes to an embedding layer, after which it is fed to an RNN model, an LSTM or a Bi-LSTM model, followed by a series of fully connected layers. The first in the series is a dense layer, after which a batch normalization layer stabilizes the input. We use the ReLU activation function, given by Eq. 7

$$ReLU = \begin{cases} 0, if\ x < 0, \\ x, if\ x \geq 0. \end{cases} \tag{7}$$

A dropout layer is used to prevent overfitting. The output thus received is flattened for dimensionality reduction. Table 6 demonstrates the input, output, and parametric information of the proposed RNN component.

Table 16: Information of each layer in the proposed RNN architecture

Layer	Input	Output	Parameters
Embedding	(None, 300)	(None, 300, 50)	50000
LSTM/Bi-LSTM	(None, 300, 50)	(None, 128)	58880
Dense	(None, 128)	(None, 256)	33024
ReLU	(None, 256)	(None, 256)	0
Dropout	(None, 256)	(None, 256)	0
Dense	(None, 256)	(None, 1)	257
ReLU	(None, 1)	(None, 1)	0

CNN Component

For image classification, CNN architectures have displayed outstanding performance in multiple domains. They are among the most popular deep architectures due to their advantage of extracting and learning implicit visual features without much pre-processing. CNNs are capable of understanding the spatial and temporal dependencies in an image which aids in better classification. CNNs are a type of neural network that performs a "convolution" operation on the input data. A convolutional operation $*$ on functions f and g is given by the formula:

$$(f * g)(t) \triangleq \int_{-\infty}^{\infty} f(\tau) g(t-\tau) d\tau \quad (8)$$

where the product of functions f and g is calculated by reversing and shifting one of these functions. The network consists of an input layer, an output layer, and multiple hidden layers. Each convolutional layer takes as its parameters the kernel size, stride and zero padding. Convolutions work with a series of pooling and fully connected layers. Feature extraction is performed using convolutional and pooling layers, where pooling layers are responsible for input dimensionality reduction. The proposed architecture uses a max-pooling operation. Fully connected layers perform the classification using sigmoid as the appropriate activation function. We use four pre-trained CNN architectures, VGG-16, InceptionV3, XceptionNet, and MobileNetV2, to fine-tune them to achieve the best performances. We also propose a simple self-designed CNN model to compare its performance with pre-existing pre-trained models. The image input is fed to a CNN model, post which bottleneck feature extraction is performed. Parameter tuning in CNN is performed similarly to RNN by adding dense, batch normalization,

ReLU, and Dropout layers. Output from the image sequence is also sent to a flatten layer. From both text and image sequences, we have received flattened outputs of the same dimension. These flattened outputs are then used to fuse the features as per the desired fusion mechanism.

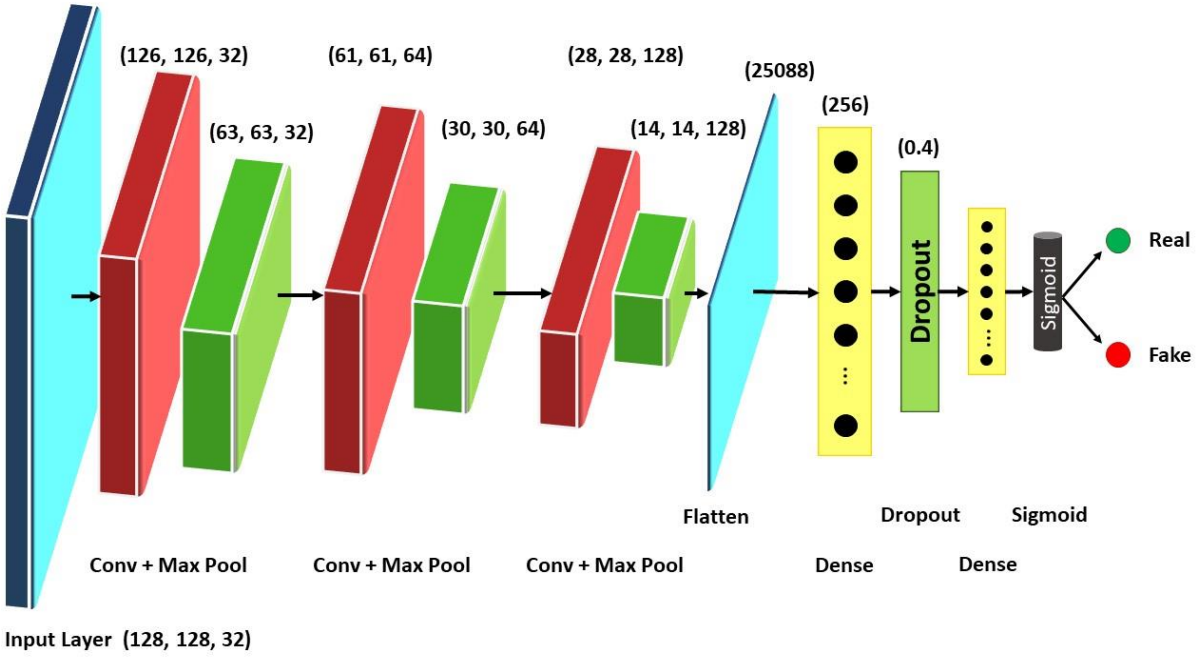


Figure 37: Proposed CNN Architecture

The proposed CNN model's architecture is represented in figure 28. This additional model has been designed to study the effect of a new convolutional model for the task, comparing the results using pre-trained models. We eliminate the separate bottleneck feature extraction stage used with other pre-trained models and let the CNN do this itself. In early fusion, the layers from CNN to Dropout in the image pipeline are replaced by this proposed CNN model, and the next flatten layer in ARCNN stays in place. To avoid redundancy, flatten layer in the proposed CNN model is removed, and layers only up to the dropout layer are added. For late fusion, this CNN architecture replaces layers starting from the CNN block to the sigmoid layer. The construct of this model contains three convolutional layers, each followed by a max pool operational layer, further flattening the feature vectors followed by fully connected layers. A dense layer is appended with which a dropout with a probability of 0.4 is used. Table 7 shows the input, output, and parametric information of the model. Proposed CNN architecture is scalable and easily reproducible.

Table 17: Information of each layer in the proposed CNN architecture

Layer	Input	Output	Parameters
Conv2D	(128, 128, 3)	(126, 126, 32)	896
MaxPooling2D	(126, 126, 32)	(63, 63, 32)	0
Conv2D	(63, 63, 32)	(61, 61, 64)	18496
MaxPooling2D	(61, 61, 64)	(30, 30, 64)	0
Conv2D	(30, 30, 64)	(28, 28, 128)	73856
MaxPooling2D	(28, 28, 128)	(14, 14, 128)	0
Flatten	(14, 14, 128)	25088	0
Dense	25088	256	6422784
Dropout	256	256	0
Dense	256	1	257

Fusion Mechanisms

An algorithmic explanation of early and late fusion variants of the ARCNN model is described by Algorithms 1 and 2. In the early fusion variant, the outputs from flattened layers are joined using simple concatenation. The next phase involves the addition of a series of dense layers with Dropout and ReLU activation functions. Classification is supported by the Sigmoid activation function for binary classification. In the Late Fusion variant, the initial phase is similar to early fusion, where text and image data are passed through RNN and CNN layers, and a sequence of fully connected layers is added, including dense, batch normalization, ReLU, and Dropout layers. Instead of flattening the outputs herein, they are led to a Sigmoid layer for individual training and generation of prediction probabilities. The classification results are obtained in the form of probabilistic values for each modality which are then combined using late fusion techniques. The fusion methodologies used in the ARCNN architecture are discussed below.

Early Fusion: In multi-modal frameworks, fusing multimedia modalities is a challenging task. Early fusion, also known as data-level fusion or fusion in feature space, combines features extracted from different data streams before training the model. Data from different streams are of different dimensions. These are to be scaled or normalized at a fixed dimension for all types of data. We have performed this using a flatten layer that brings features to the same scale. Feature vectors V_t and V_i from different data streams are integrated into a single large vector V_c. This combined vector handles all multi-modal features and performs a one-time training. The combination of vectors is carried out by an operation between V_t and V_i which is a simple concatenation operation in our case. The operation represents it:

$$V_c = V_t \oplus V_i \quad (9)$$

where $\oplus$ is the operator between the two. Early fusion is an advantageous approach as it learns features in a collaborative environment as a unified representation of data streams. No separate training phases are required for each data stream. Features from all data streams are combined, and then a single training phase is carried out. It makes the process faster and efficient. Figure 29 shows the flow of the early fusion process.

Algorithm 1: Early Fusion with ARCNN

Input: $A = \{a_1, a_2, \ldots a_n\}$ is set of text vectors, $B = \{b_1, b_2, \ldots b_n\}$ is set of images of size 128*128, $Y = \{y_1, y_2, \ldots y_n\}$ is a set of labels for A and B.

1. Split A, B, and Y into three subsets as $\{(A_1, B_1, Y_1), (A_2, B_2, Y_2), (A_3, B_3, Y_3)\}$ for 60% training, 20% validation and 20% testing.
2. **For** $i = 1$ to 3, **do**
3. Add respective RNN and CNN models $M_1$and M_2.
4. Extract bottleneck features from M_2 for image input B.
5. Append series of fully connected layers (dense, batch normalization, relu) to $M_1$and M_2.
6. Apply dropouts with 0.5 probability to $M_1$and M_2.
7. Flatten both text and image feature vectors thus obtained, V_t and V_i. to make them unidimensional.
8. Combine V_t and V_i using concatenation and obtain a combined feature vector, $V_c =, V_t \oplus V_i$
9. Add fully connected layers (dense, batch normalization, relu) to V_c setting dropout values as 0.4.
10. Apply binary sigmoid classifier and calculate final prediction probabilities P_f.
11. Calculate the performance of the testing set.
12. **end for**
13. **Return** performance on the testing set.

Algorithm 2: Late Fusion with ARCNN

Input: $A = \{a_1, a_2, \ldots a_n\}$ is a set of text vectors, $B = \{b_1, b_2, \ldots b_n\}$ is a set of images of size 128*128, $Y = \{y_1, y_2, \ldots y_n\}$ is a set of labels for A and B.

1. Split A, B, and Y into three subsets as $\{(A_1, B_1, Y_1), (A_2, B_2, Y_2), (A_3, B_3, Y_3)\}$ for 60% training, 20% validation and 20% testing.
2. **For** $i = 1$ to 3, **do**
3. Add respective RNN and CNN models $M_1$and M_2.
4. Extract bottleneck features from M_2 for image input B.
5. Append series of fully connected layers (dense, batch normalization, relu) to $M_1$and M_2.
6. Apply dropouts with 0.5 probability to $M_1$and M_2.
7. Add binary sigmoid classifier to both $M_1$and M_2 individually and obtain independent prediction probabilities P_t and P_i on testing set
8. Combine P_t and P_i using late fusion operations and obtain combined prediction probabilities, $P_c = P_t \odot P_i$.

Late fusion operations:

$$P_c = P_{av} = (P_t + P_i)/2$$
$$P_c = P_s = P_t + P_i$$
$$P_c = P_m = max(P_t, P_i)$$
$$P_c = P_w = P_t * w_t + P_i * w_i$$

9. Calculate the performances on the testing set.
10. **end for**
11. **Return** performances on the testing set.

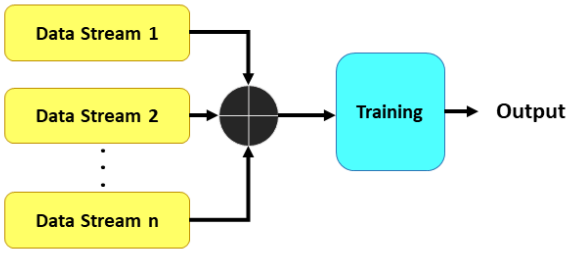


Figure 38: Early Fusion Systematic Flow

Late Fusion: Also known as decision level fusion, late fusion is performed later based on the classification decisions from all data streams. Late fusion is easier to perform and provides a simple and scalable architecture. Learning of features is performed before integration, whereas, in early fusion, features are combined first and then passed for training. Each data stream of different modalities is fed to a training model, and decisions are extracted in terms of prediction probabilities. These prediction vectors are then combined using a suitable combinatorial operation. Figure 30 represents the flow of the late fusion process. We use decision level scores from text and image stream in the proposed work and fuse them accordingly.

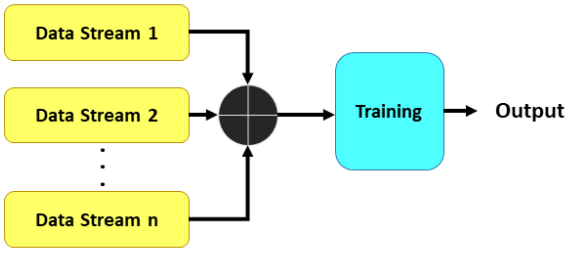


Figure 39: Late Fusion Systematic Flow

The fusion function f that fuses decisions of text and image streams is denoted by Eq. $f : P_t, P_i \rightarrow P_c$ where P_t *and* P_i are two different feature maps that denote the decisions of each stream in probabilistic values. The combined probabilities denoted by P_c gives the output decisions after late fusion. The late fusion scores thus obtained are denoted as P_{av} (Average), P_m (Maximum), P_s (Sum) and P_w (Weighted Average).

Average Fusion: Average fusion combines modalities by taking a simple average of prediction vectors. Mathematically, it is represented as:

$$P_{av} = (P_t + P_i)/2 \quad (20)$$

where combined prediction P_{av} is calculated by averaging, i.e., summing up values from all streams and then dividing by the number of data streams. Combining only text and image features, the number of data streams is 2, which divides the sum of P_t and P_i.

Max Fusion: This technique uses the maximum value of probability, i.e., prioritizing the decision with a larger weight or value than the other to select the higher contributing score between the feature maps. This is performed by a simple maximum function denoted as:

$$P_m = max(P_t, P_i) \tag{31}$$

Sum Fusion: Sum fusion sums up the values of feature maps obtained from both data streams simply by adding up the values. It is expressed as

$$P_s = P_t + P_i. \tag{42}$$

Weighted Average Fusion: In this fusion mechanism, we assign random weights w_t and w_t to feature maps from both streams. This has an advantage over the other methods as it helps decide which data type contributes to better detection. Playing with the values of assigned weights provides a route to experimentation to decide which weights would make the model best performing. Mathematically, the arbitrary weights, ranging from 0.0 to 1.0, each weight complimenting the other, are multiplied by their respective prediction probability arrays and then summed up. It is defined as

$$P_w = (P_t * w_t + P_i * w_i) \tag{53}$$

where w_t and w_i are weights assigned for text stream and image stream respectively.

4.4 EXPERIMENTAL RESULT ANALYSIS

Datasets

For the performance evaluation of the proposed ARCNN framework, we use six multi-modal datasets that include news articles and tweets containing real and fake information. Among these, we created two datasets, two subsets of the ReCOVery dataset containing a collection of news articles and associated tweets, CoAID dataset with health-related tweets, and MediaEval 2020 benchmark dataset. We evaluate the two subsets of the ReCOVery dataset separately to analyze the effect of corpora on the performance of our model. Detailed information on these datasets is provided as follows in table 8:

Table 18: Details of datasets used

Dataset	Referred to as	Type	Real News Count	Fake News Count	Total
CovID I	D1	Articles, posts	1059	1310	2369
CovID II	D2	Tweets, posts	1171	1303	2474
ReCOVery [141]	D3	Articles	1345	651	1996
ReCOVery [141]	D4	Tweets	3968	924	4892
CoAID [142]	D5	Tweets	565	517	1082
MediaEval [143]	D6	Tweets	791	289	1080

CovID I

We introduce this dataset in lieu of the urgent need for infodemic datasets to suffice the requirement of deep learning algorithms. It is a multi-modal dataset consisting of textual and visual information of fake news related to coronavirus. CovID I consist of fake and real news articles from websites and social media posts. This dataset is referred to as D1 in the results section. The sources of fake news are various fact-checking websites registered with Poynter IFCN. True news has been extracted from official news website articles.

CovID II

This dataset has been proposed to assist the study of the effect of corpora on the proposed ARCNN model. Fake news originates in both social media posts and fake news articles. As our detection is primely knowledge-based, the writing styles of text have an impact on the detection. Distinguishing text based on how a sentence is written, its formation, vocabulary, and grammar play a significant role in our task. There is a difference in the ways social media posts are written from the way official news is structured. It is essential to differentiate between true and real social media posts since both follow a writing style different from news articles. We used a mix of true articles for this detection, primarily extracted from Twitter posts and a few from news articles, to create an unbiased set with a mix of fake posts and articles. This dataset has been referred to as D2.

ReCOVery

Zhou et al. introduced this multi-modal repository consisting of 2029 news articles on COVID-19 collected between January to May 2020 containing textual, visual, temporal, and network information [141]. 140820 tweets related to these news articles are also added to the dataset. We utilize these news articles and tweet ids as separate datasets, hereafter referred to

as D3 and D4, respectively, for textual-visual detection. Items containing both textual and visual information are only used, and the rest are discarded.

CoAID

Cui and Lee proposed a COVID-19 Healthcare Misinformation Dataset , a repository of health-related fake news spread via news websites and on Twitter [142]. The dataset contains news article titles, user tweets, and associated user interactions, i.e., tweet replies. Since the image URLs were not available for news articles, we utilized only the tweet IDs available in the dataset to extract multi-modal information. We were finally left with 565 real and 517 fake tweets containing both textual and visual content.

MediaEval 2020

The onset of the COVID-19 pandemic coincided with the release of 5G technology which gave rise to an entirely distinguished conspiracy that claimed that the arrival of COVID-19 was due to the masts of 5G networks. This led to a violent situation of destroying 5G poles in the UK. MediaEval 2020 issued a benchmark dataset for fake news detection, which is a collection of misinformation related to 5G linked coronavirus conspiracies, other COVID conspiracies, and non-conspiracy tweets [143]. We categorize all the conspiracy tweets within a single label of fake news and used non-conspiracy tweets as real tweets.

Implementation Settings

All experiments have been performed on Google Colab which provides up to 13.53 free RAM and 12 GB NVIDIA Tesla K80 GPU. The proposed framework is built and implemented in Python 3 using Keras deep learning framework. Input data is split into 60% training, 20% validation, and 20% testing. All models have been trained with binary cross-entropy for 15 epochs with a batch size of 64. In the late fusion variant of ARCNN, where image and text models are trained separately, we have used Adam and RMSprop optimizers for image and text classifiers, respectively. In the early fusion variant, we have used Adam optimizer.

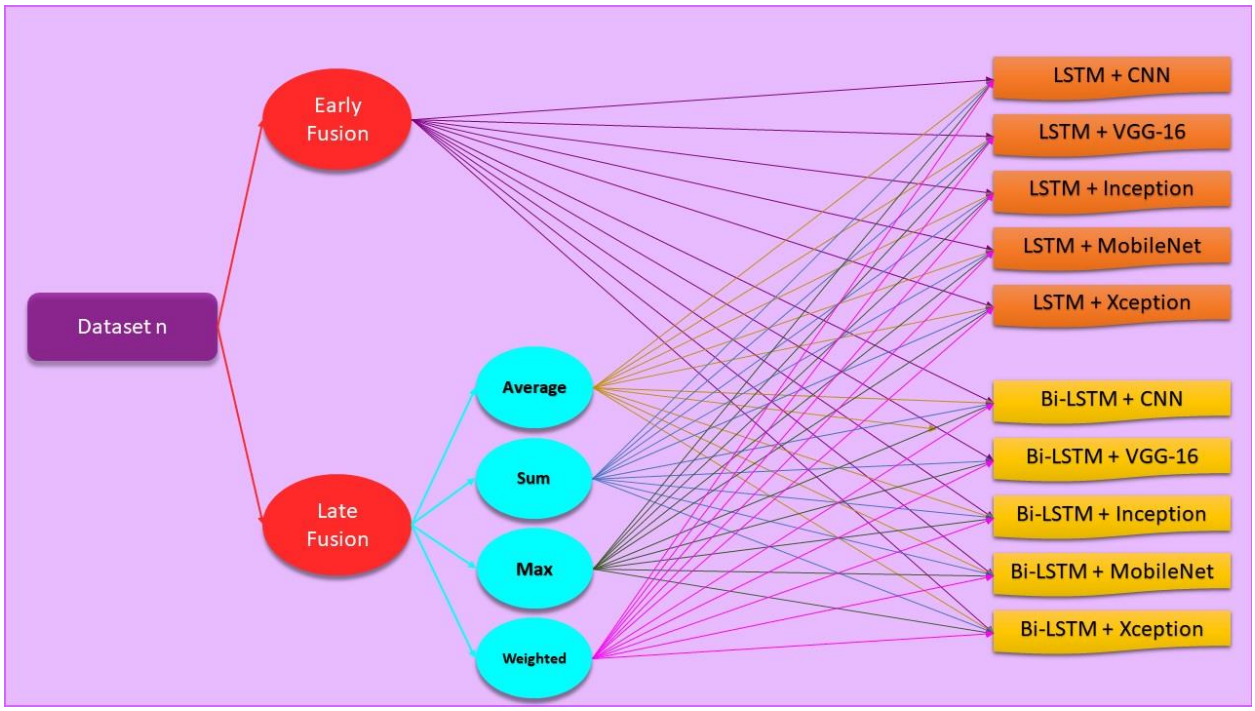


Figure 40: Various combinations of classification models and fusion methods used for experimentation

Table 19: RNN and CNN models used for text and image classification and their combinations

Model	RNN (Text)	CNN (Image)	Combination
M1	LSTM	CNN	LSTM + CNN
M2		VGG-16	LSTM + VGG-16
M3		InceptionV3	LSTM + InceptionV3
M4		MobileNetV2	LSTM + MobileNetV2
M5		XceptionNet	LSTM + XceptionNet
M6	Bi-LSTM	CNN	LSTM + CNN
M7		VGG-16	LSTM + VGG-16
M8		InceptionV3	LSTM + InceptionV3
M9		MobileNetV2	LSTM + MobileNetV2
M1		XceptionNet	LSTM + XceptionNet

Training is performed on ten different combinations of RNN and CNN models. Early fusion models are trained and evaluated separately as they follow a different training route than late fusion. Late fusion models on all datasets are run separately, evaluating all late fusion methods on a single training and testing run for each dataset. Thus, for one dataset evaluating one model, we extract five sets of results (one set belonging to one fusion method), thus corresponding to each of the ten combinations of RNN and CNN models. We obtain a total of 50 sets of results for each dataset. A description of model settings is provided in table 9.

We have employed a wide range of evaluation metrics for performance comparison of the proposed framework. The performance scores are listed in F1-measure, accuracy, precision, recall (TPR), FPR, ROC, specificity, and MCC score. These values can be calculated using confusion matrix values as described by the mathematical equations that follow:

$$F1\ score = \frac{2TP}{2TP+FP+FN} \tag{14}$$

$$Accuracy = \frac{TP+TN}{P+N} \tag{15}$$

$$Precision = \frac{TP}{TP+FP} \tag{16}$$

$$Recall\ (True\ Positive\ Rate) = \frac{TP}{TP+FN} \tag{17}$$

$$False\ Positive\ Rate\ (FPR) = \frac{FP}{FP+TN} \tag{18}$$

$$Specificity = \frac{TN}{FP+TN} \tag{19}$$

$$Mathew's\ Correlation\ Coefficient\ (MCC) = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)*(TP+FN)*(TN+FP)*(TN+FN)}} \tag{20}$$

Results

This section presents the results obtained by performing all the experiments on six datasets. The consolidated results are presented in table 10 for each dataset. On each dataset, we use ten model combinations, M1 to M10, and the fusion on each one of them is performed in five different ways, thus summing up to 50 experiments on each dataset.

Result Analysis

From the wide set of results, the highest accuracies obtained in each dataset are represented in figure 32. D2 (CovID II) and D5 (CoAID) demonstrate their top accuracy values as 100%. This shows that all the news items in the testing sets were classified correctly by models M1 (LSTM + CNN) and M10 (Bi-LSTM + Xception), respectively. Both of these datasets include a majority of social media posts, D2 with a collection of posts from various online social networks, and D5 with a collection of tweets. The remaining datasets have also achieved good classification accuracies with different models. Comparing the F1 scores for each of the six datasets in figure 33, we observe that D2 and D5 obtain 100% and 98.54% scores, representing a balanced dataset. F1 scores come out to be good when the dataset has balanced items for each category. Slightly lower accuracies or F1-scores are also accounted for

the imbalanced nature of datasets used and model selection. Overall, the proposed architecture has provided good results for fake news classification.

Table 20: Accuracy percentage of proposed ARCNN on six datasets

	Fusion Mechanisms	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
D1	Early Fusion	83.43	89.43	86.29	90.29	86.57	82.86	83.43	80.29	92.86	86.57
	Avg Fusion	82.96	88.86	83.14	92.00	88.57	87.48	88.29	82.29	87.14	86.29
	Max Fusion	75.32	82.29	76.86	81.43	80.29	78.18	84.57	80.86	83.43	83.43
	Sum Fusion	75.09	82.29	76.86	81.43	80.29	77.22	84.57	80.86	83.43	83.43
	Weighted Avg	88.85	89.43	84.86	92.00	88.86	90.05	89.43	87.71	89.43	88.29
D2	Early Fusion	100.0	93.55	99.46	96.51	98.66	62.54	96.51	77.69	88.98	98.92
	Avg Fusion	99.51	86.29	87.63	90.59	91.94	99.51	85.22	83.06	88.98	87.63
	Max Fusion	83.83	84.95	86.83	89.52	90.86	83.09	86.56	87.10	90.86	89.78
	Sum Fusion	83.83	84.95	86.83	89.52	90.86	83.09	86.56	87.10	90.86	89.78
	Weighted Avg	99.76	98.39	98.12	98.39	98.39	100.0	100.0	100.0	100.0	100.0
D3	Early Fusion	80.12	76.66	79.25	77.81	78.67	81.66	75.50	80.98	77.10	77.52
	Avg Fusion	80.12	79.48	74.28	73.70	79.48	81.66	75.79	74.64	78.96	77.81
	Max Fusion	75.68	78.90	73.99	73.99	73.99	77.03	78.39	75.22	79.54	78.96
	Sum Fusion	75.87	78.61	73.99	73.99	73.99	76.83	78.39	75.22	79.54	78.96
	Weighted Avg	80.12	83.82	73.7	76.88	79.77	81.66	78.10	74.64	80.98	79.54
D4	Early Fusion	82.31	91.72	92.43	92.22	92.02	85.17	92.73	91.5	91.91	91.91
	Avg Fusion	81.47	91.40	90.07	92.43	90.07	84.64	91.72	86.90	91.61	91.20
	Max Fusion	76.35	90.99	90.89	90.89	90.48	81.36	92.23	91.91	92.12	91.81
	Sum Fusion	76.39	90.99	90.89	90.89	90.48	80.99	92.23	91.91	92.12	91.81
	Weighted Avg	82.31	92.73	92.22	92.43	92.02	85.17	91.72	91.50	91.91	91.91
D5	Early Fusion	87.35	80.09	81.02	63.43	85.19	87.35	89.81	88.43	84.72	91.20
	Avg Fusion	97.85	80.65	76.96	84.33	78.80	95.38	78.34	78.80	84.79	79.23
	Max Fusion	88.00	89.86	88.94	92.63	92.52	83.08	91.71	89.86	92.63	89.86
	Sum Fusion	87.69	89.86	88.94	92.63	92.52	82.46	91.71	89.86	92.63	89.86
	Weighted Avg	98.46	97.70	97.70	98.62	97.24	97.23	94.93	96.77	95.39	94.93
D6	Early Fusion	84.81	80.57	84.36	84.36	84.36	85.44	85.78	85.78	86.73	86.26
	Avg Fusion	85.76	67.30	84.36	81.99	84.83	86.71	67.30	83.41	83.89	84.83
	Max Fusion	86.08	84.83	84.36	84.36	84.36	86.08	84.36	84.36	84.36	84.83
	Sum Fusion	86.08	84.83	84.36	84.36	84.36	86.08	84.36	84.36	84.36	84.83
	Weighted Avg	86.71	84.83	85.31	84.83	84.83	86.39	84.83	86.26	84.83	85.78

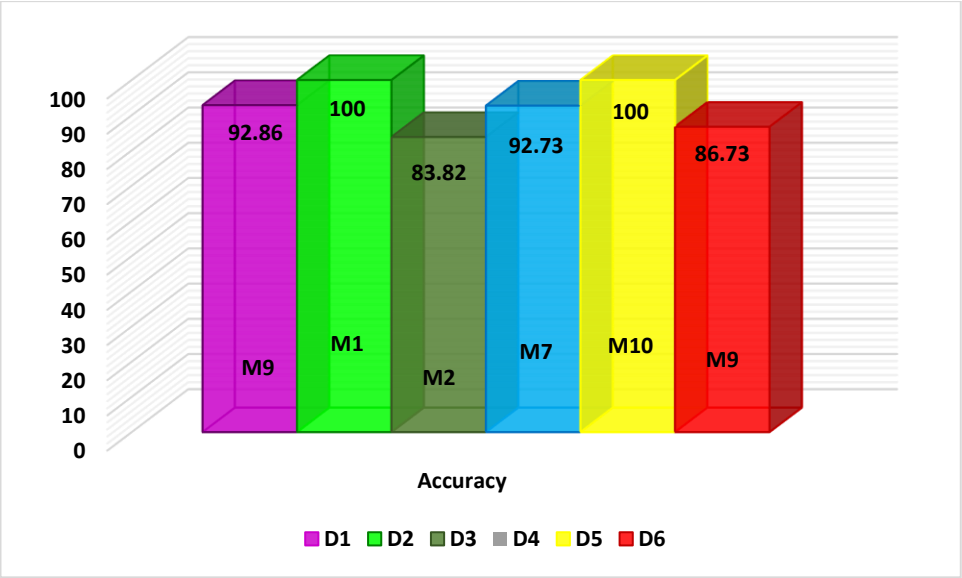


Figure 41: Highest accuracies obtained in each dataset

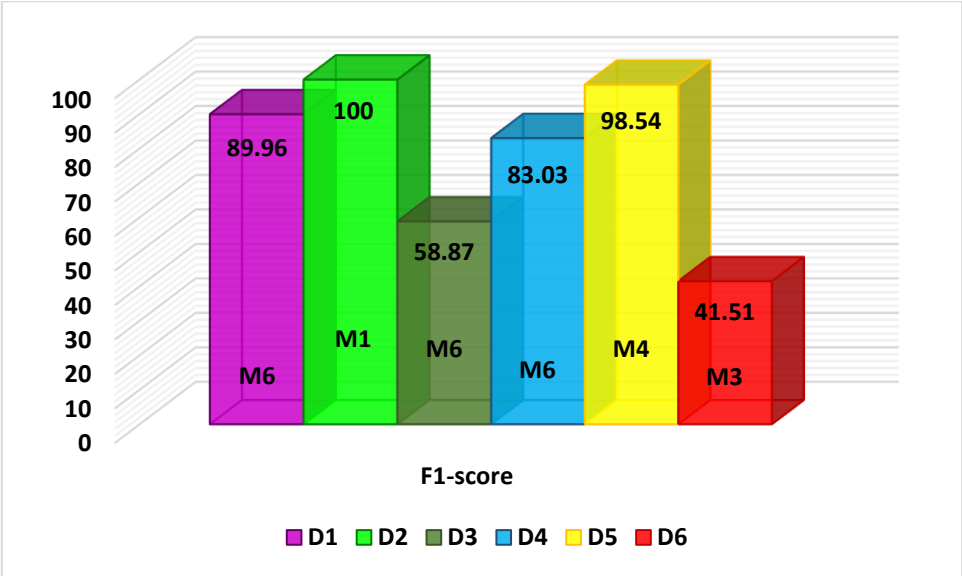


Figure 42: Highest F1-scores obtained in each dataset

Performance Comparison on Each Dataset based on Classification Model and Fusion Method

This section analyses the experimentation results using ten classification models, each with five fusion methods. The graphs represent the accuracy trends for the same. Observation from these trends suggests that high accuracies are obtained using weighted average fusion and early fusion. The highest accuracies on all datasets tend to lie between the range of 80% to 100%. This indicates that the proposed ARCNN model is an effective multi-modal classifier. The figures allow us to understand and select the best performing classification models and fusion methods for such tasks. Despite receiving good performance from all models, it is worth noting that models using Bi-LSTM display marginally better results. However, both LSTM and

Bi-LSTM have performed equally well. In terms of image classifiers, VGG-16, MobileNetV2, and proposed CNN models are observed to provide the highest results for all datasets. According to experimental observation, VGG-16 took a longer duration of training time than other models. This adds up as a disadvantage for VGG-16, despite providing good results. Other classifiers are comparatively faster acceptably good classification accuracies.

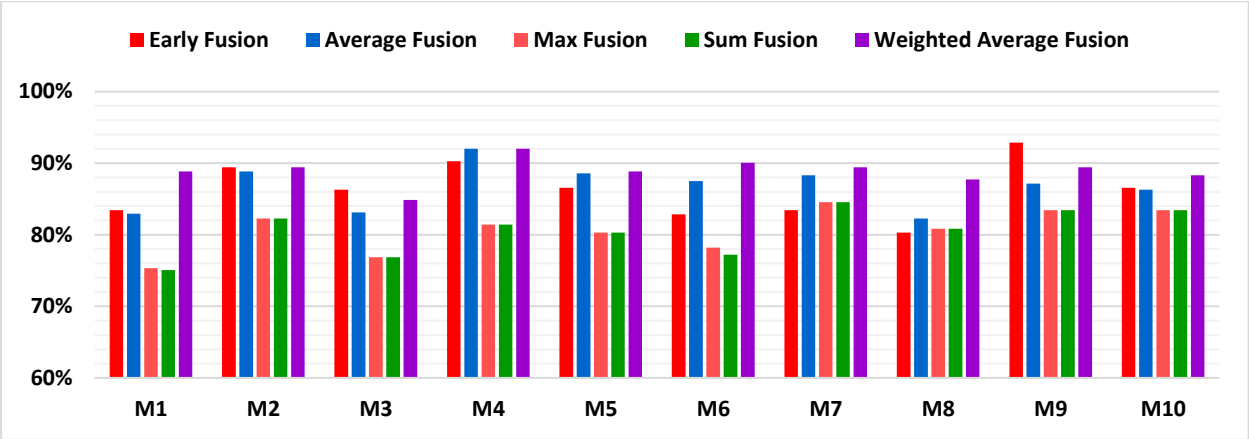


Figure 43: Performance Comparison on D1

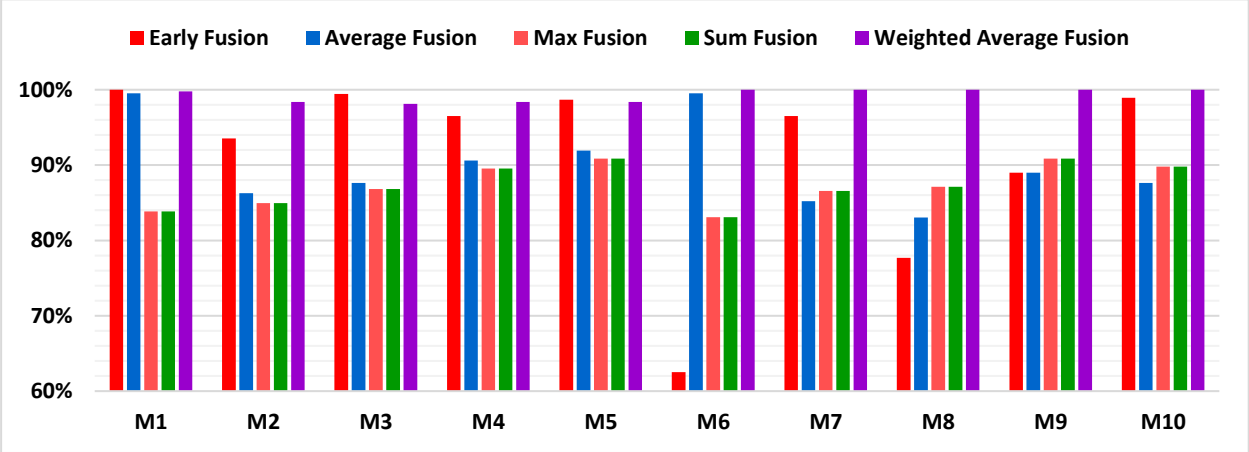


Figure 44: Performance Comparison on D2

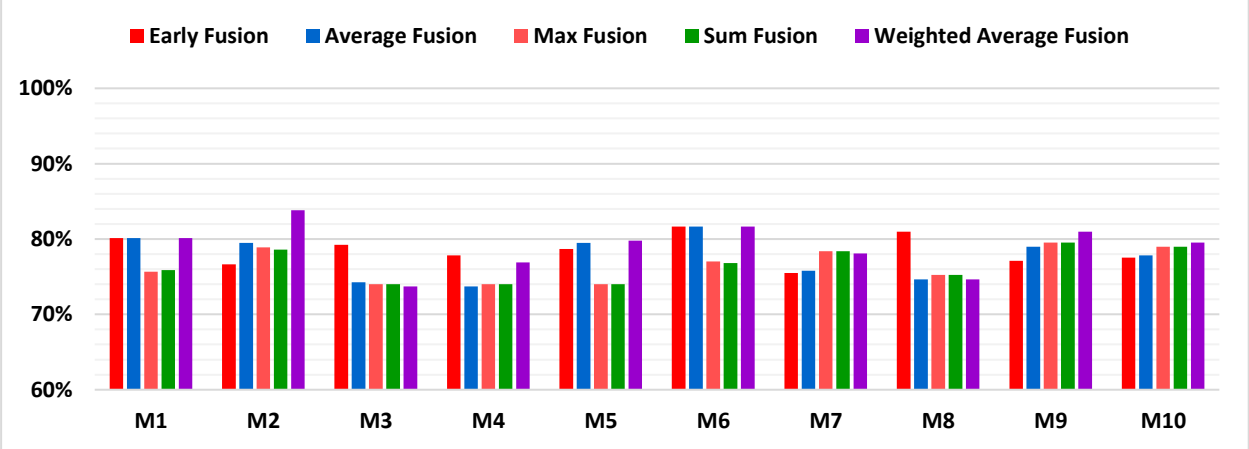


Figure 45: Performance Comparison on D3

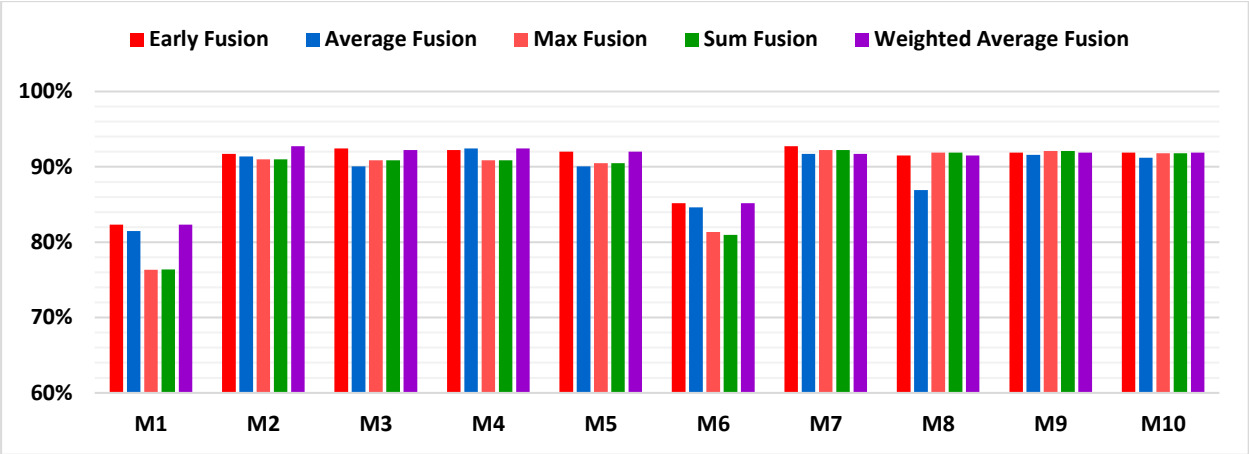


Figure 46: Performance Comparison on D4

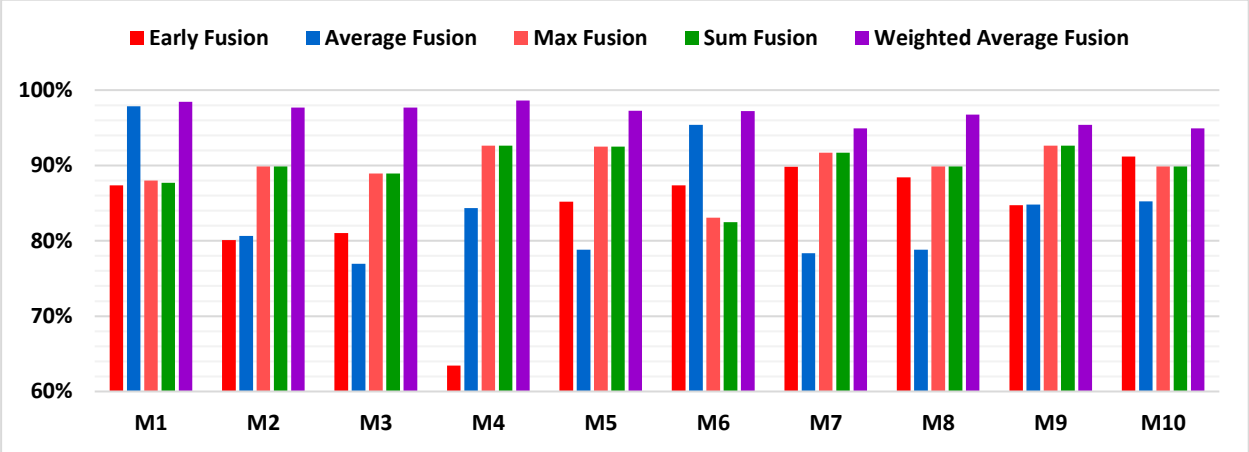


Figure 47: Performance Comparison on D5

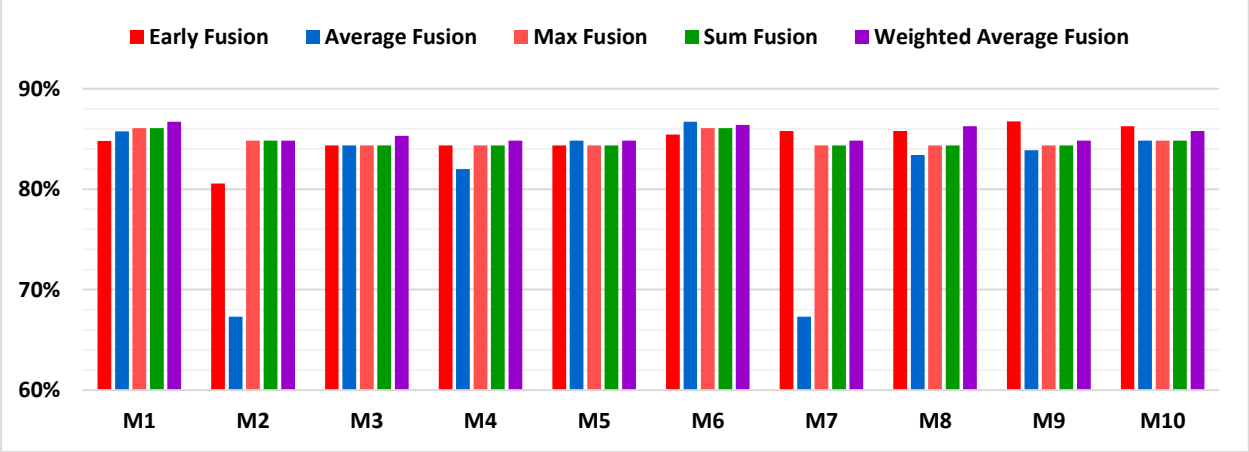


Figure 48: Performance Comparison on D6

Comparative Analysis of Classification Models on Six Datasets

Further, we narrow down our analysis to interpret the consistency in the performance of the ten classification model combinations used in our experiments. We observe that although Model M9 (Bi-LSTM + MobileNetV2) achieved the highest accuracy on dataset D1, Model

M4 performed more consistently as the average accuracy is much closer to the maximum accuracy. Model M2, despite delivering more consistent results, falls below the average and maximum accuracy values of M4, respectively. In the D2 dataset, M1 seems an obvious winner, followed by M5 and M4 models. D3 dataset, flooded with news articles of long and complex texts, poses a credible challenge to the performance of all ten fusion techniques. The Maximum and Average accuracy is lower than that obtained in all other datasets for most of the fusion methods. Most fusion methods perform equally well in dataset D4, with high repeatability and consistency in giving accurate results. In dataset D5, the model continuously learns to differentiate between a fake and real piece of information, and hence most of the methods provide highly accurate detection, similar to dataset D2. This can be attributed to the fact that there is a balance in Real and Fake news count proportion in these datasets. D6, a highly biased dataset regarding the number of Real news counts, has the most recurring troughs (M2 and M7) in the plot. None of the methods provides significant results (above 90%) as they do with other datasets. The outcome is observed to be indifferent irrespective of the method chosen. We can observe that the classification models have achieved good maximum and average accuracies for most datasets. Datasets achieving slightly lower results (~80%) are attributed to the type of information. We observe that our image classification models have performed well overall. Lower than 90% accuracy scores in D3 are observed due to complex and lengthy texts of articles. D6 being a highly biased dataset produces accuracy scores close to 85%.

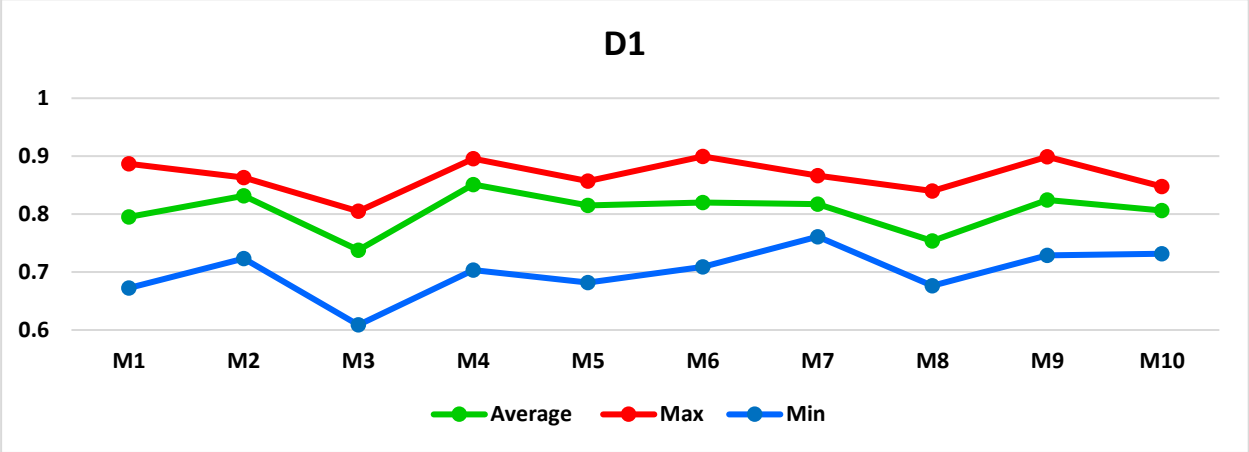


Figure 49: Comparative Analysis of Classification Models on D1

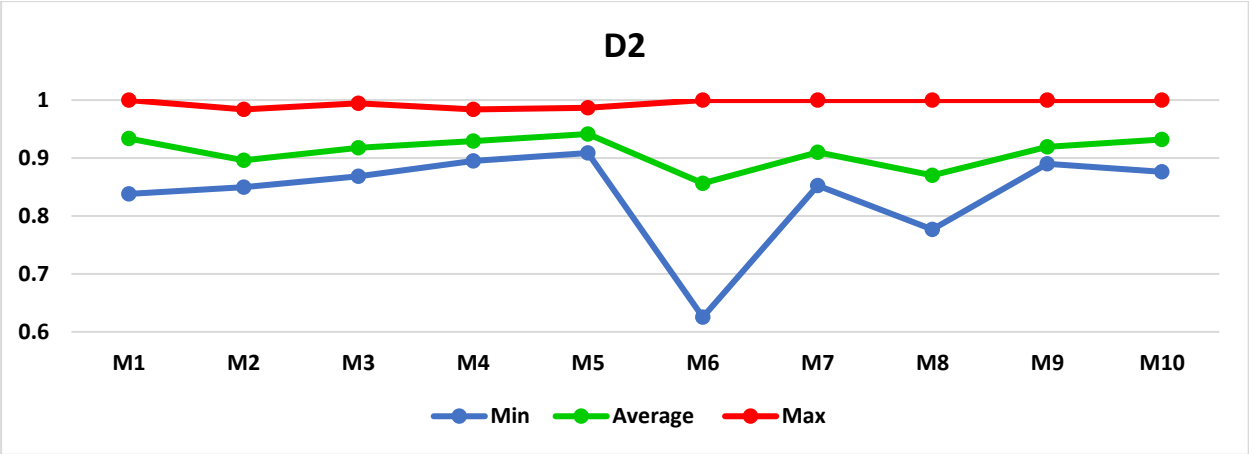


Figure 50: Comparative Analysis of Classification Models on D2

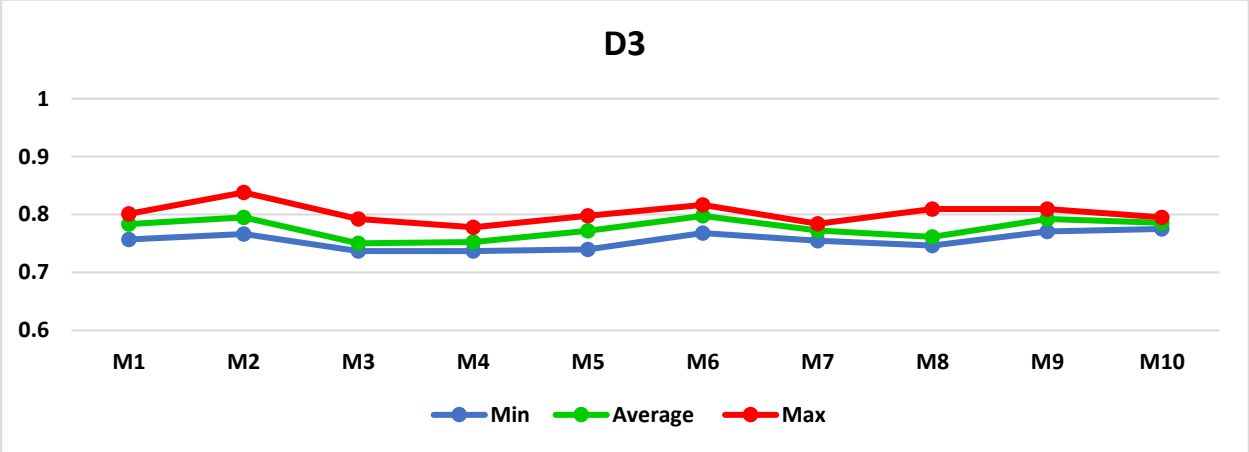


Figure 51: Comparative Analysis of Classification Models on D3

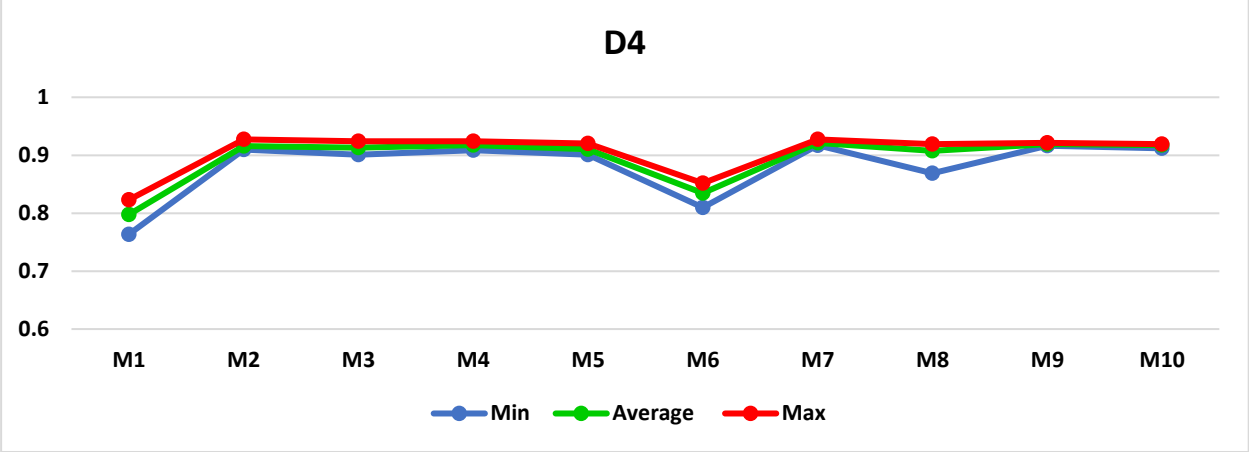


Figure 52: Comparative Analysis of Classification Models on D4

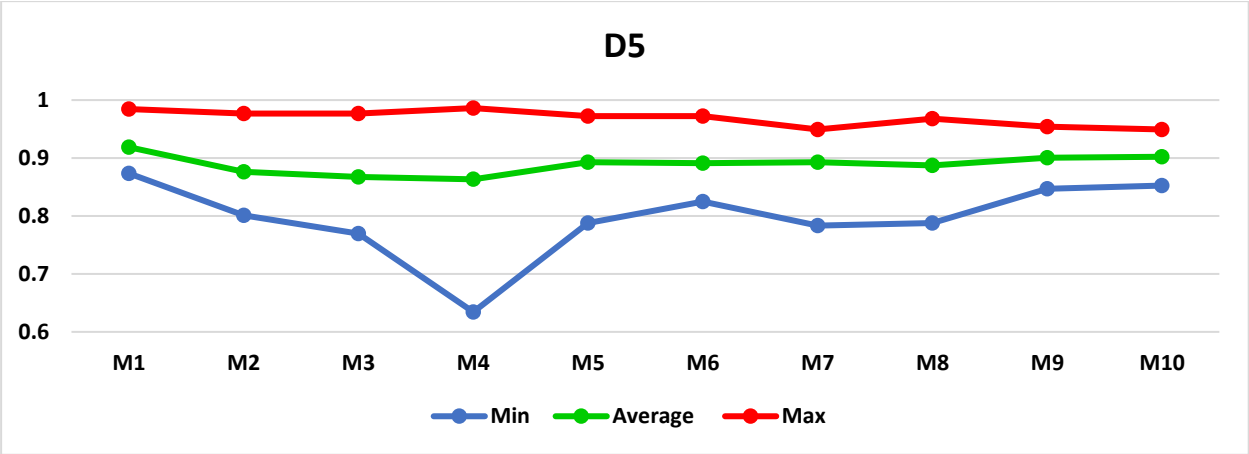


Figure 53: Comparative Analysis of Classification Models on D5

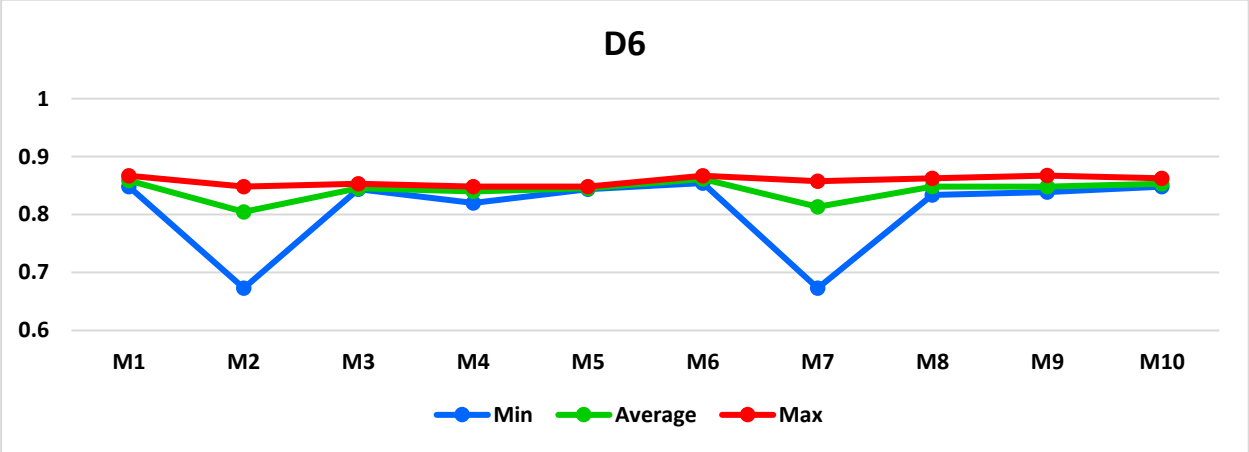


Figure 54: Comparative Analysis of Classification Models on D6

Comparative Analysis of Classification Models on All Fusion Methods

Another important criterion that decides the accuracy of fake news detection is the fusion technique applied to each model. The first of the five techniques applied is early fusion. Models M1, M3, M5, and M10 perform significantly better than others in early fusion. In the remaining four late fusion techniques, we observe that M1 and M5 perform consistently better or at par when compared to the other eight models. M1, where LSTM is used with proposed CNN architecture, produces consistent results over all the ten models and all fusion techniques. Hence, we can state that the proposed CNN architecture offers excellent stability in classification on all datasets. M5 with LSTM and XceptionNet is also observed to produce favorable results under all circumstances.

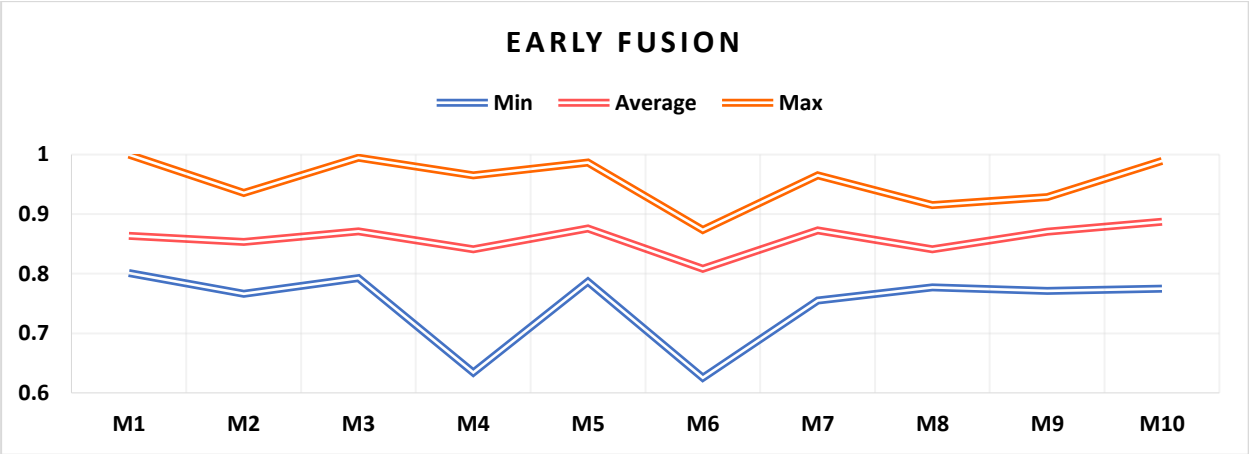


Figure 55: Comparative Analysis of Early Fusion with all Classification Methods

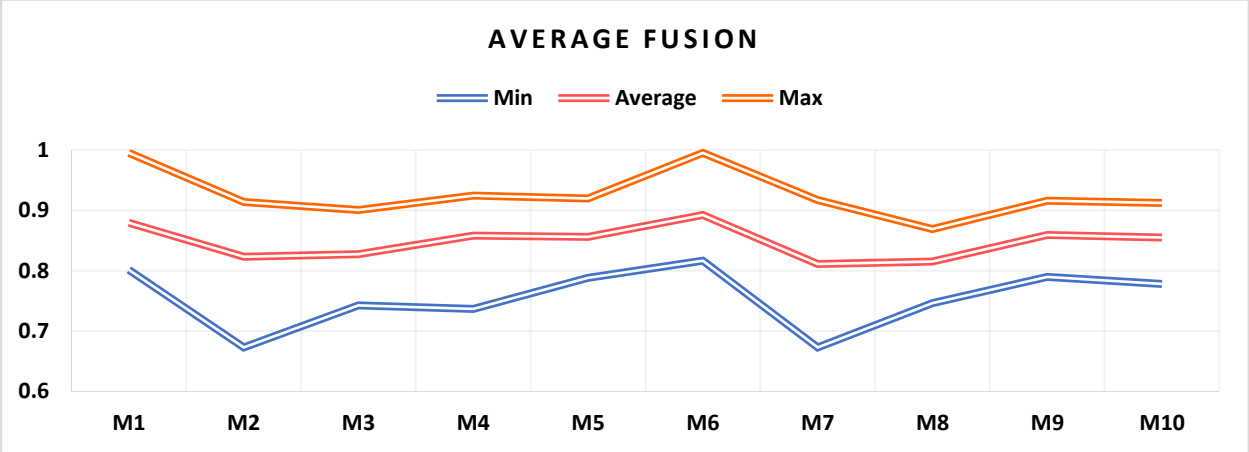


Figure 56: Comparative Analysis of Average Fusion with all Classification Methods

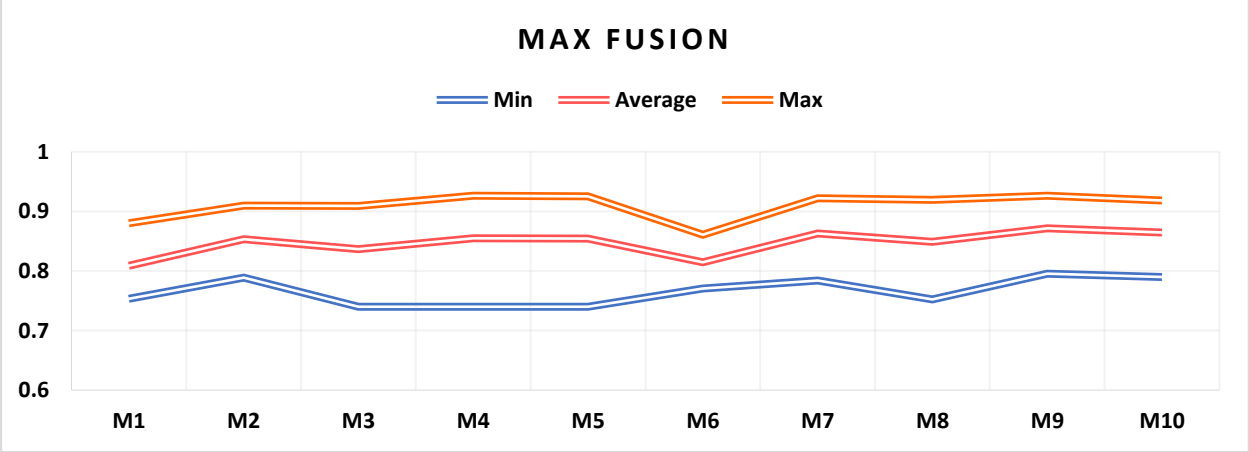


Figure 57: Comparative Analysis of Max Fusion with all Classification Methods

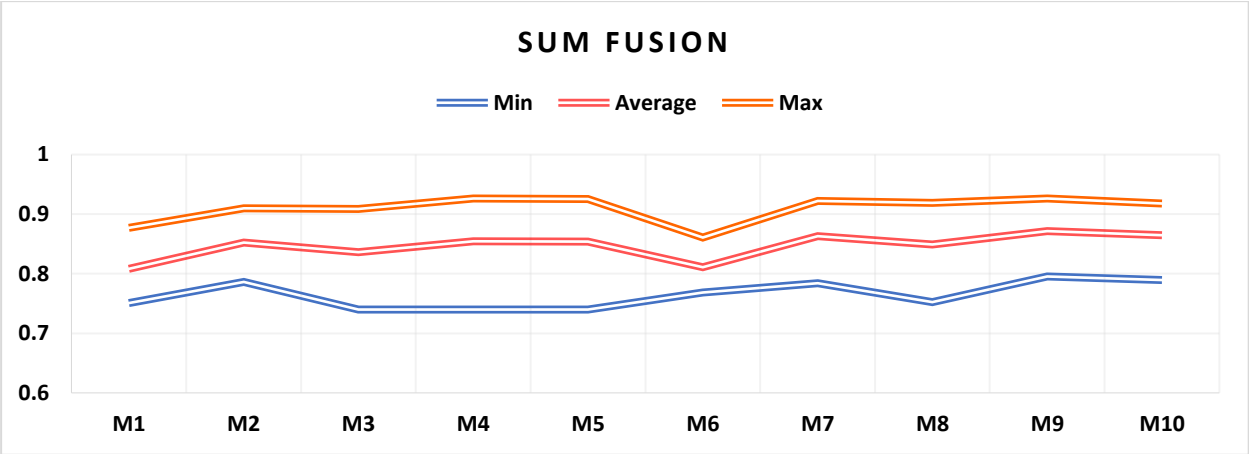


Figure 58: Comparative Analysis of Sum Fusion with all Classification Methods

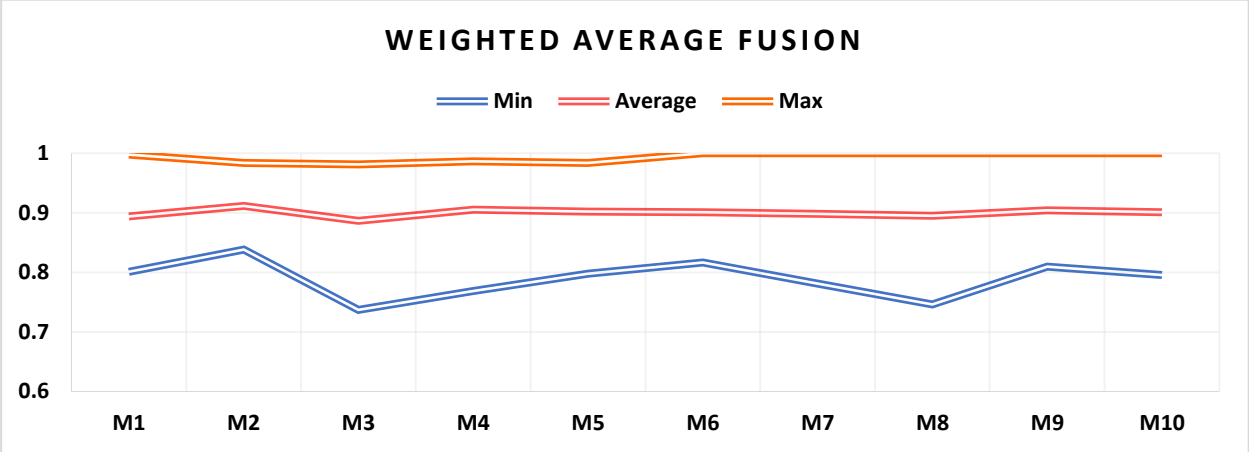


Figure 59: Comparative Analysis of Weighted Average Fusion with all Classification Methods

Comparative Analysis of Fusion Methods on Six Datasets

Observing the performance trends of all fusion methods for a dataset, this analysis is performed to select the best fusion methods. Below are presented fusion-based comparative graphs for each dataset. In D1, we see the best maximum and average performance with weighted average fusion. The next best performers are early and average fusions. Max and sum fusion methods have performed lower than all other methods. Their maximum fusion results are often lower than average results from other fusion methods. In D2, we observe that the maximum accuracy of 100% is portrayed by early fusion and weighted average fusion. Average fusion stands third in terms of maximum and average accuracies. Comparing overall performance, weighted average fusion is a stable performer with fewer deviations among minimum, average and maximum results. Whereas we view early fusion performance in D2, there is a considerable gap or deviation among the three. Observing the trends in dataset D3, we can state that all fusion methods have performed at a similar scale. Maximum accuracy

values are greater than or equal to 80% in each case. Providing the highest max and average scores, weighted average fusion stands the best with dataset D3 also. In D4, maximum accuracies for all fusions are over 90%, and average values are also close to 90%. Early fusion and weighted average fusion have given the best results, while average fusion stands next. Max and Sum fusion methods have also shown promising results on this dataset. In D5, we see that average fusion is next best to the weighted average, with early fusion standing third in terms of maximum results, but there is more deviation in both of them than weighted-average fusion. D6 achieves average and quite similar performance using all fusion methods, with only average fusion fluctuating more among the three categories. An overall analysis of fusion methods considering all datasets can conclude that weighted average fusion displays the highest performance with the slightest variance. Early fusion stands next in terms of performance and lower variance. Average fusion is a good performer but might demonstrate instability among maximum, average, and minimum results. Sum and Max fusion have often provided similar results, both showing lower metrics in all cases.

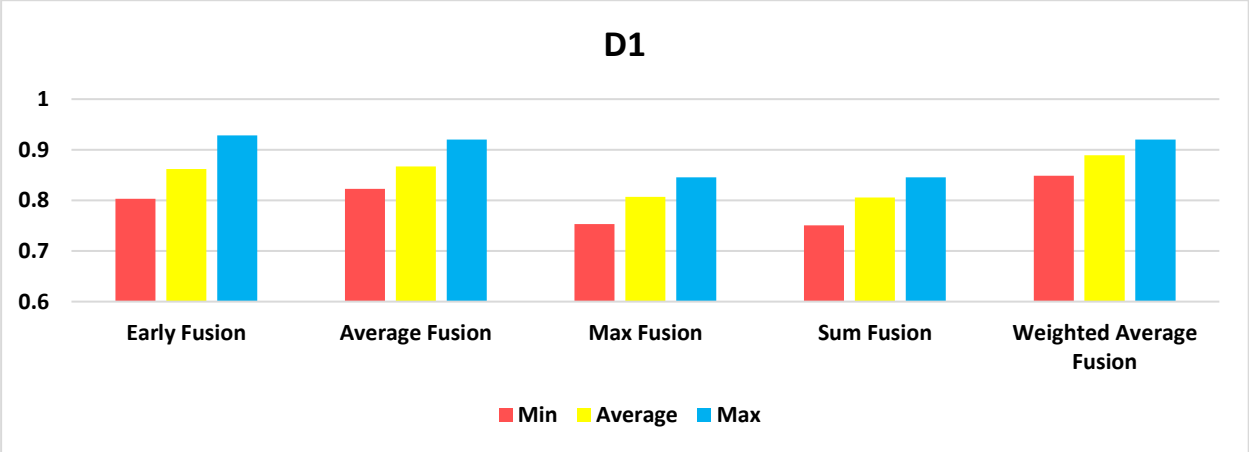


Figure 60: Comparative Analysis of Fusion Methods on D1

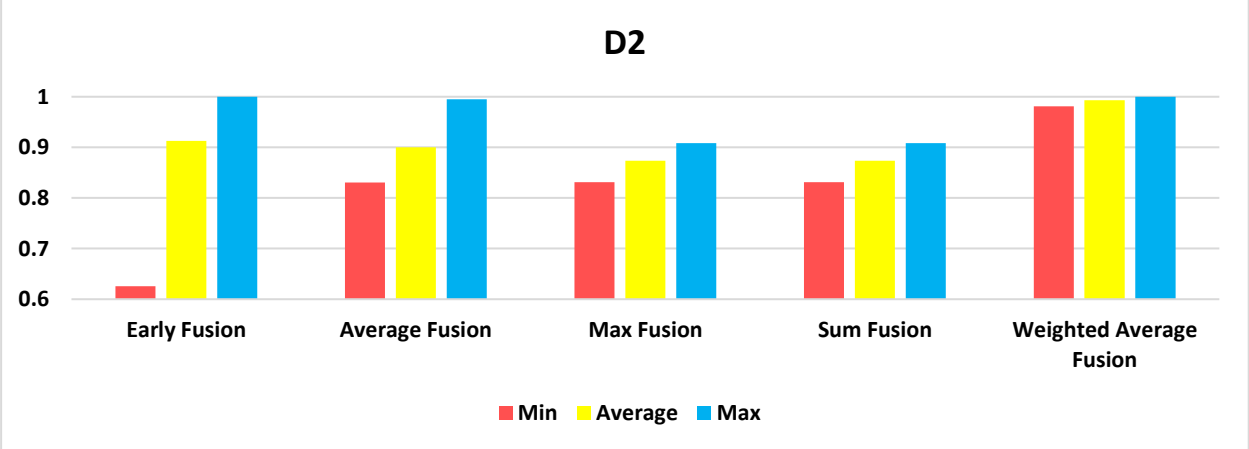


Figure 61: Comparative Analysis of Fusion Methods on D2

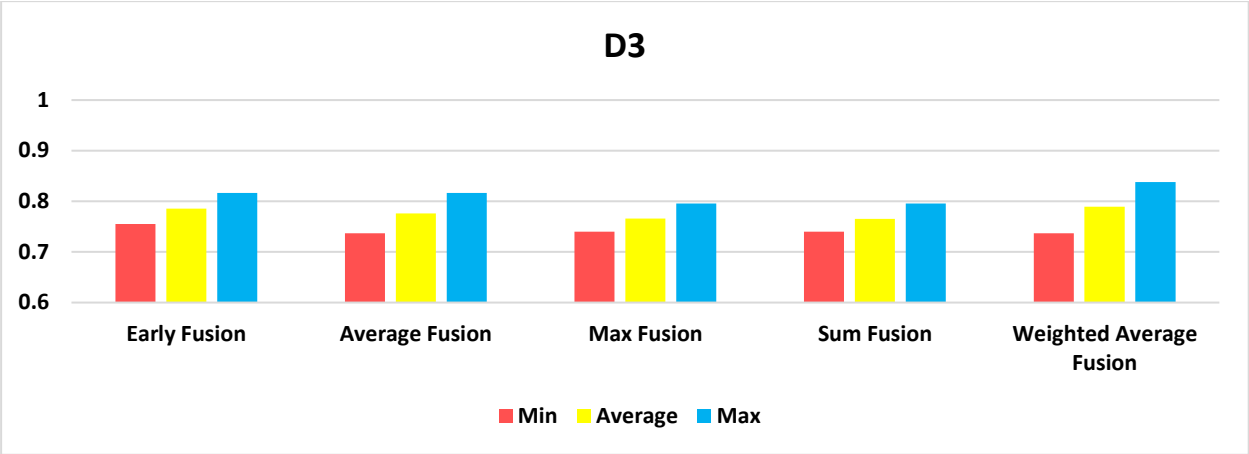


Figure 62: Comparative Analysis of Fusion Methods on D3

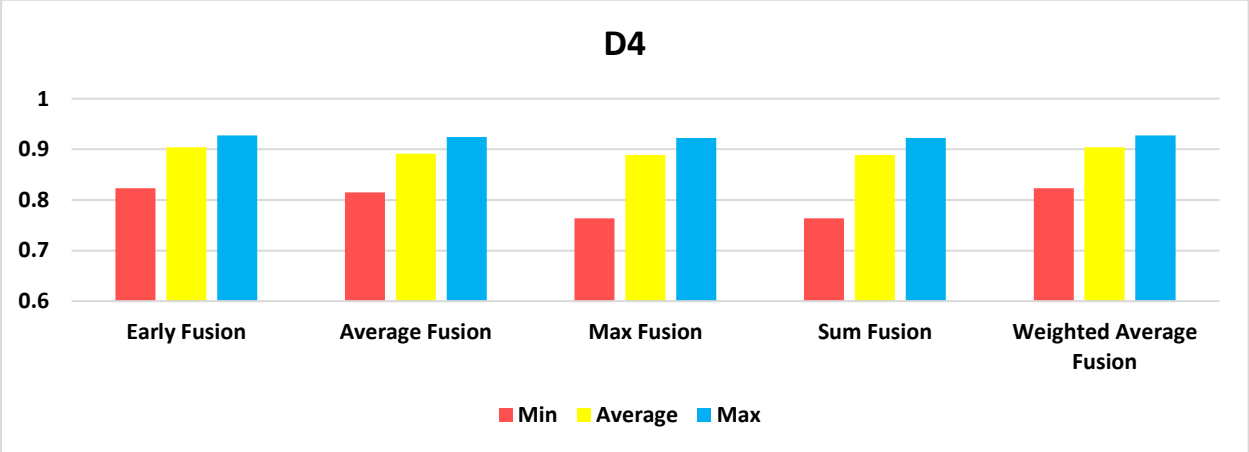


Figure 63: Comparative Analysis of Fusion Methods on D4

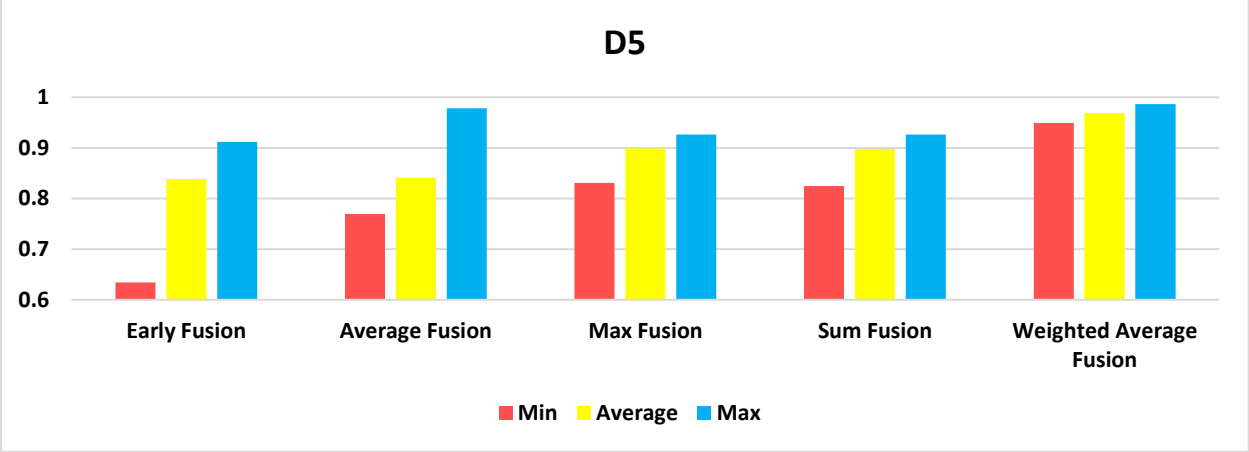


Figure 64: Comparative Analysis of Fusion Methods on D5

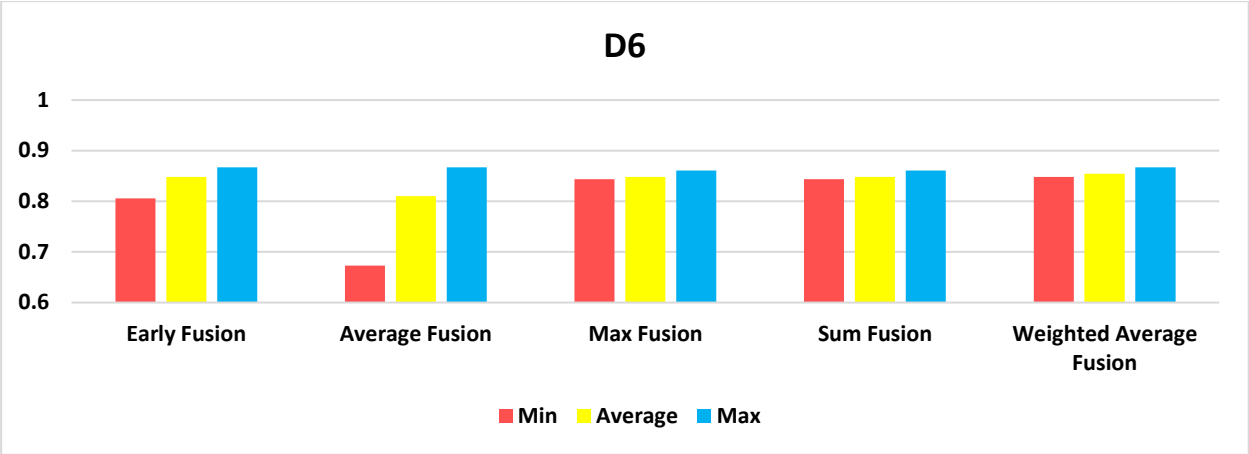


Figure 65: Comparative Analysis of Fusion Methods on D6

Text and Image Contribution in Weighted Average Fusion

In weighted average fusion, weights are assigned to each data modality, where weights can be described as their contribution to the final prediction results. Each modality, text, and image is assigned a value between 0 to 1 to represent their share in the final results. We experimented with values to generate optimum results with the weighted average fusion technique, alternating among the weights. The results mentioned in this work are the best results obtained under such fusion. We demonstrate the weights assigned to the modalities in each classification model for datasets D1 to D6. Analyzing the text and image contribution individually for each dataset, we observe multiple trends. This variation is attributed to the quality of data present in the datasets. In D1, most classification models work well by assigning 65% weightage to text and the remaining 35% to images. Observations in D2 are quite varying where four models are used with 85% text and 15% image weightage. Rest models are used with 30-40% image weightage and the remaining to text. Dataset D3 works well by assigning 40-50% weightage to visual data. In D4, values fluctuate, displaying most stability with 35-45% for the image. In D5, most models work best by assigning 35% to the image, while others use 25% weightage for the image. In D6, weight assigning for achieving the highest possible accuracies has shown a random trend for all models. Overall, we demonstrate that visual data is a compelling factor in fake news detection, with their average contribution between 30% to 50% when combined with textual modality.

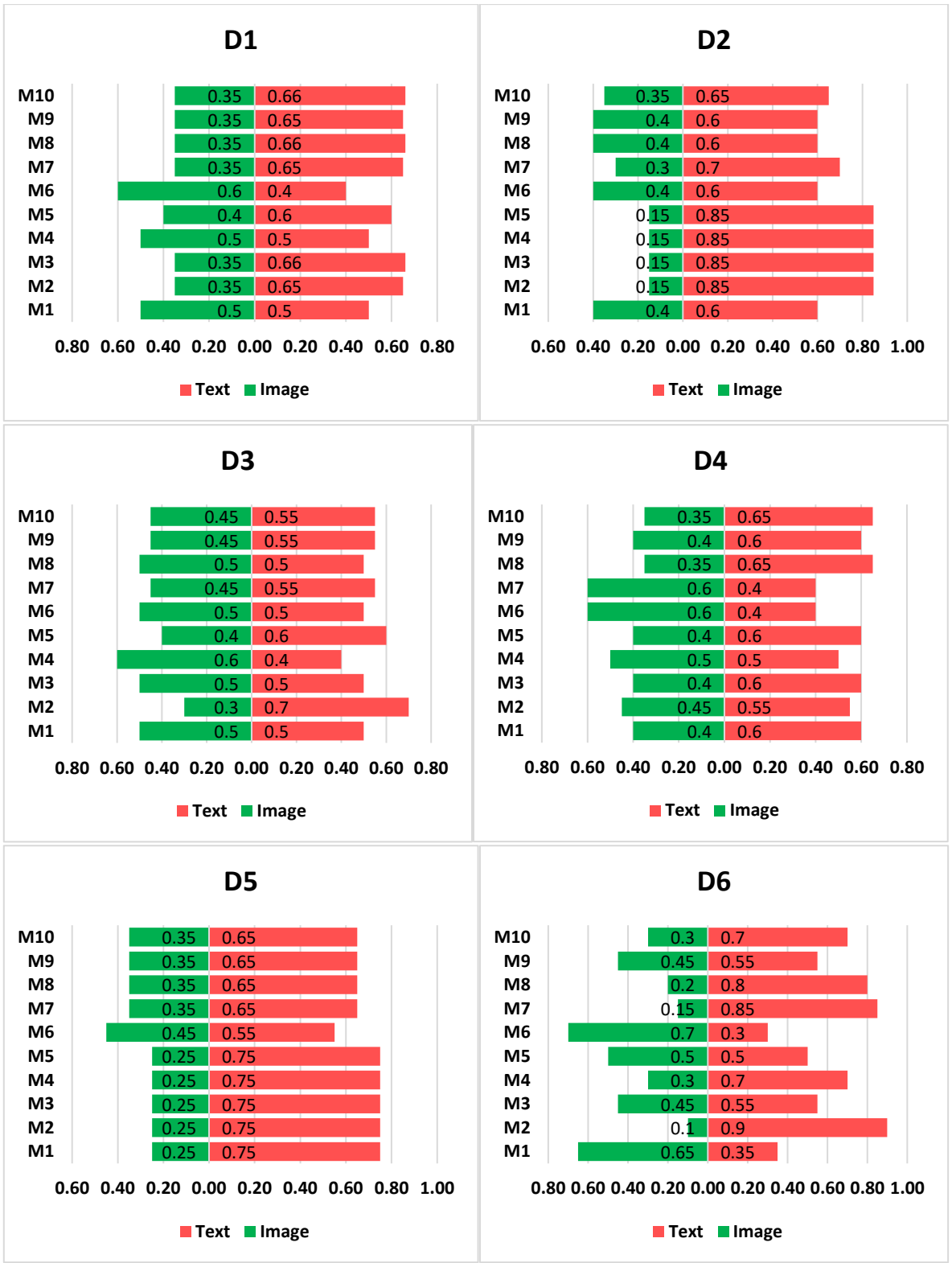


Figure 66: Text and Image contributions in all datasets for weighted average fusion

Overall Performance Analysis

To conclude, we provide comparisons based on overall performances by deciding optimum classification models and fusion methods. This establishes Weighted Average Fusion

as the best fusion method with the highest maximum, average, and minimum results. The next best performance is shown by early fusion and average fusion, but their average and mins are less significant than the weighted average. Ranking all the fusion methods according to their performance, weighted average fusion stands apart as the best, with early fusion being second and average fusion as the third. Max fusion and Sum fusion appear to be at the same level with similar results.

Observing the trends obtained by classification models M1 to M10, the highest maximum performance (=100%) is shown by M1, M6, M7, M8, M9, and M10. Models M6 to M10 use Bi-LSTM for text classification, which makes a better choice than LSTM. Analyzing the averages we get, XceptionNet and MobileNetV2 make the best choice for image classification on the datasets we have used. Also, considering the minimum results provided by each model, M10 and M9 lead succeeding M1, M3, M5, and M8. Therefore, we can say that the proposed ARCNN model with specified hyperparameters gives excellent fake news detection provided by any pre-trained classification models.

We see that different ranges of result scores are obtained for a different dataset. For datasets containing tweets as text, scores are higher as compared to datasets with news articles. This is because tweets are short statements, whereas articles are long and complex posts. Accuracy scores are also different for different sizes of datasets. Datasets in which classes are balanced offer more effective predictions than unbalanced datasets. The size and quality of corpora play a significant part in building an effective classification mechanism.

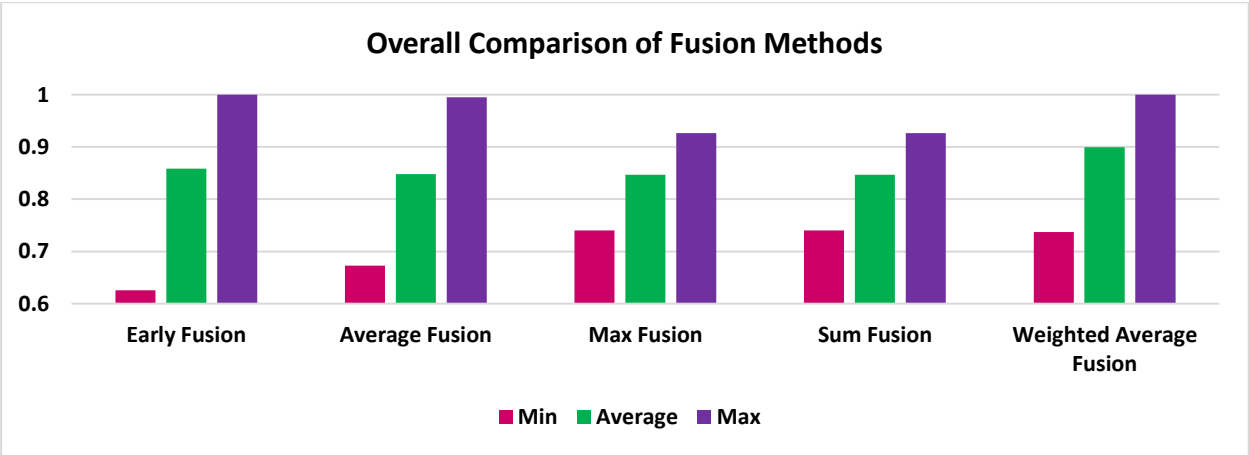


Figure 67: Overall performance comparison of fusion methods

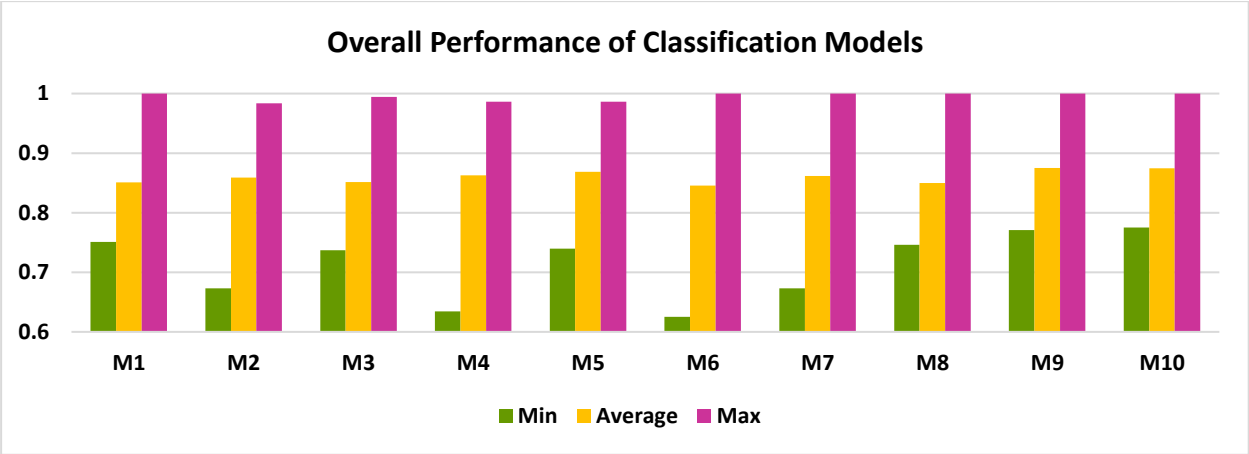


Figure 68: Overall performance comparison of classification models used

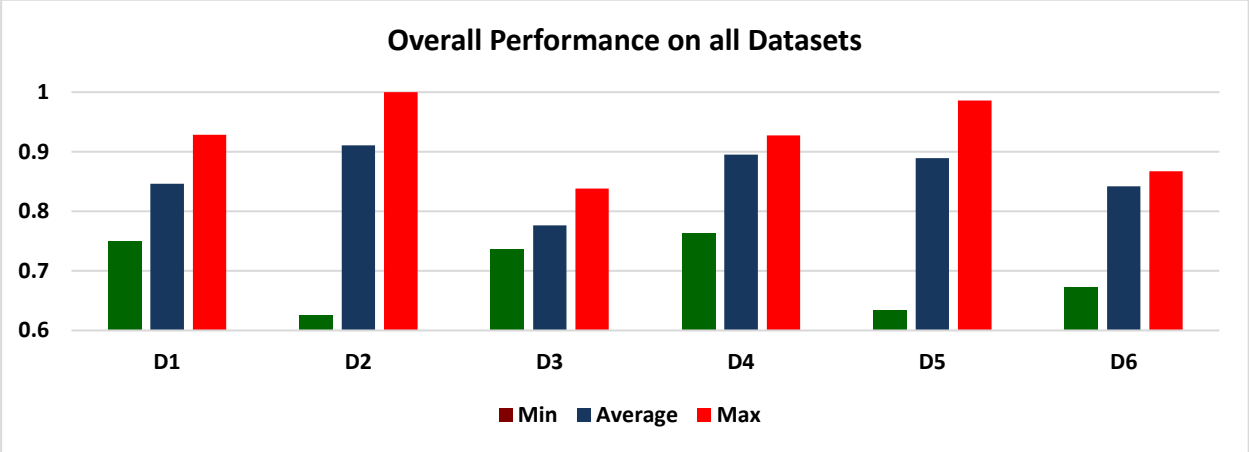


Figure 69: Overall performance comparison on all datasets`

Ablation Study

Ablation study is the procedure of systematic framework analysis by the removal of its components. It helps in identifying the usefulness of each component of the framework separately. We perform the ablation study to examine the contribution of text classification and image classification models. We experiment with the individual techniques, LSTM, Bi-LSTM, Proposed CNN, VGG-16, InceptionV3, MobileNetV2, and XceptionNet on the six real-world datasets. The parameter settings are kept identical to that of the overall ARCNN framework. Table 11 reports the accuracy percentage of the ablation study on six datasets. The last row in the table illustrates the highest performance observed by the ARCNN framework wherein text and image components are combined.

Table 21: Ablation study of proposed ARCNN framework

Feature	Accuracy (%)					
	D1	D2	D3	D4	D5	D6
Individual Techniques						
Text-based Techniques						
LSTM	79.43	96.24	81.51	89.67	96.29	83.71
Bi-LSTM	82.56	98.32	81.92	91.33	96.61	84.01
Image-based Techniques						
Proposed CNN	79.98	90.64	65.37	87.14	89.03	76.54
VGG-16	87.83	92.65	67.82	89.33	92.54	79.03
InceptionV3	87.91	95.81	74.12	90.12	96.82	84.61
MobileNetV2	89.25	90.73	73.46	88.65	96.54	83.02
XceptionNet	88.44	96.06	77.10	91.31	91.46	81.09
Overall ARCNN Framework						
Text + Image	92.86	100.00	83.82	92.73	98.62	86.73

Baseline Comparison

To authenticate the worthiness of the proposed model, we compare our approach with existing baselines on six datasets in terms of accuracy scores. We observe that the proposed approach performs better than the existing ones, as depicted in table 12. We perform a comparison based on single modality models and multi-modality approaches:

Single Modality Models

The proposed work uses a combinatorial approach of predictions based on both text and image modalities. We also analyze the results obtained by individual text and image classifiers. For both of these, fake news detection is carried out using features obtained from only one data type.

Text: Being more advantageous and with better performing results, we used a Bi-LSTM layer for textual feature extraction from all datasets. The layer is followed by a dense layer, batch normalization, ReLU activation, and Dropout of value 0.4. Classification is supported by sigmoid activation with RMSprop optimizer and binary cross-entropy.

Image: Displaying good results over ARCNN, we decide on using our proposed CNN model for unimodal classification. The description of the model has been provided in 4.3. Images from the datasets are input to the model to classify fake news only depending on implicit features of the visual information provided.

Multi-modality Models

In the multi-modal scenario, we compare our approach with three established baselines. Since these methods have not been analyzed on covid-specific fake news datasets, we reproduce their works by providing their models with a similar setup as their own. All experiments are performed by training these models on all six datasets, and results are evaluated.

Att-RNN: Jin et al. have proposed a fusion architecture that incorporates textual, visual, and social features and combines them using an attention mechanism [144]. For a fair comparison, we combine only textual and visual features. As proposed in their approach, we have used LSTM for text and VGG-19 pre-trained on the Imagenet dataset for the images.

The hidden layer dimension for text is set to 32, and the tanh activation function is used. The entire network is trained for 100 epochs with early stopping with a batch size of 128.

EANN: In this approach, the Text-CNN model is used for textual feature extraction, and VGG-19 is used for visuals [145]. Features from both streams are concatenated as an early fusion to form a single set of feature maps, and the model is trained thereafter. We eliminate the event discriminator used in EANN. We train the model for 100 epochs using early stopping and a batch size of 64.

TI-CNN: Yang et al. utilizes implicit and explicit text and image features and then combine them with early fusion [146]. We use implicit features pre-existing in text and image and train the model after early fusion. The textual branch consists of one-dimensional convolution, while the visual branch uses three-dimensional convolutional layers. Features from both CNNs are joined using concatenation, and the model is trained for 100 epochs with early stopping with a batch size of 64.

Table 22: Baseline Comparison

Dataset	Method	Accuracy	Precision	Recall	F1-Score
D1	Textual	0.7654	0.7352	0.7718	0.753
	Visual	0.663	0.6312	0.6614	0.6459
	Att-RNN [144]	0.7308	0.25	0.6667	0.3636
	EANN [145]	0.8132	0.6301	0.8679	0.7302
	TI-CNN [146]	0.8522	0.7815	0.9163	0.8435
	ARCNN	**0.9286**	**0.8409**	**0.9652**	**0.8988**
D2	Textual	0.9645	0.9469	0.98	0.9631
	Visual	0.876	0.8249	0.9125	0.8665
	Att-RNN [144]	0.8303	0.7807	0.8588	0.8179
	EANN [145]	0.8534	0.7379	0.9157	0.8172
	TI-CNN [146]	0.9676	0.9595	0.9708	0.9651
	ARCNN	**1**	**1**	**1**	**1**
D3	Textual	0.6548	0.3974	0.7111	0.5254
	Visual	0.6458	0.2699	0.688	0.4584
	Att-RNN [144]	0.6792	0.3701	0.7484	0.5204
	EANN [145]	0.7424	0.3721	0.6957	0.4848
	TI-CNN [146]	0.801	0.4102	0.8439	0.4871
	ARCNN	**0.8382**	**0.44**	**0.7021**	**0.541**
D4	Textual	0.7901	0.4906	0.7879	0.6047
	Visual	0.7608	0.5243	0.7619	0.7426
	Att-RNN [144]	0.7458	0.5124	0.7511	0.7013
	EANN [145]	0.7885	0.5747	0.834	0.7333
	TI-CNN [146]	0.8294	0.5978	0.8528	0.7401
	ARCNN	**0.9273**	**0.618**	**0.9735**	**0.756**
D5	Textual	0.7828	0.7459	0.8528	0.7958
	Visual	0.6926	0.6447	0.7686	0.7012
	Att-RNN [144]	0.7628	0.722	0.7401	0.7766
	EANN [145]	0.8229	0.7097	0.7841	0.7952
	TI-CNN [146]	0.8976	0.797	0.7981	0.8118
	ARCNN	**1**	**0.8526**	**0.8182**	**0.8351**
D6	Textual	0.6112	0.6089	0.6754	0.1616
	Visual	0.5561	0.0026	0.1250	0.0263
	Att-RNN [144]	0.7355	0.1052	0.7882	0.1826
	EANN [145]	0.7825	0.0826	0.7236	0.1726
	TI-CNN [146]	0.8030	0.1025	0.7902	0.1926
	ARCNN	**0.8578**	**0.1212**	**0.8**	**0.2105**

The high accuracies in our proposed framework are due to the hyper-parameter settings, model fine-tuning and fusion mechanisms. The model architectures have a high impact on classification results. For text classification, baselines have used LSTM and Text-CNN models and for image classification, VGG-19 and CNN has been used. The proposed ARCNN uses

novel architectures of RNN and CNN models. The models are designed after several experimentations of parametric settings and fine-tuning to provide best results. The differences in model architectures owe to the highly accurate classification results. The results received through sum and max fusion are quite similar to the baseline models as in the Att-RNN, EANN and TI-CNN models. We observe that highest accuracies in the proposed method are achieved by the weighted average fusion. The assignment of appropriate weights to text and image models results in high performance.

4.5 SUMMARY

In this work, we have proposed the ARCNN architecture for fake news detection. Our framework uses ten combinations of text and image classification models to detect fake news based on two modalities: text and image. We provide a generic architecture that can incorporate a pre-trained classification model of choice. We used LSTM and Bi-LSTM in the RNN component of the framework for text-classfication. In the CNN component, we used the proposed CNN, fine-tuned VGG-16, MobileNetV2, InceptionV3 and XceptionNet for image classification. We have conducted experiments on six COVID-19 fake news datasets alternating with various text and image classification models. Our introduced datasets, CovID I and CovID II, have been publicly made available. The two streams of data are combined using early fusion and four types of late fusion techniques. We have presented vast experimentation and study in fake news detection. Proposed architecture outperforms various state-of-the-art fake news detection models. Results have been calculated in terms of eight evaluation metrics for all conducted experiments. To facilitate an easy understanding of results, the data has been neatly represented using various graphs. Trends have been observed and analyzed for fellow researchers providing a deep study that can be readily utilized to build fake news detection models. Our work leverages deep learning and combines various techniques to develop a novel and scalable fake news detection mechanism. In the present scenario, there is a lack of infodemic datasets and detection mechanisms. This puts forward the challenge of distinguishing fake news from real, and hence, dealing with it becomes problematic. Coronavirus-related fake news datasets are yet limited to textual information. Datasets containing various other information like visual data or meta-data, which could serve helpful in detection, are very few. We have proposed an architecture that uses two modalities. Our proposed architecture is flexible in accepting more data streams from different modalities that can be fused. Due to the less availability of versatile data, we have exploited text and images used in social media posts and news articles. We intend to utilize video features to serve

detection based on fake news videos. We also encourage fellow researchers to build a holistic fake news detection framework that could capture most of the possible details in a piece of news and exploit them for efficient fake news detection. Future works also include the building of our proposed framework as an application or browser plugin. Collection of versatile and balanced real-world datasets and designing better mechanisms to detect fake news in real-time is promoted.

CHAPTER 5

BERT-MULTISCALE CNN FRAMEWORK

5.1 PROPOSED METHODOLOGY

We propose a novel framework using BERT and Multiscale CNN to address the fake news issue. The idea behind the framework is to build a multi-modal algorithm that identifies false information based on linguistic and visual features present in a microblog post. In this section, we elaborate on the components of the proposed framework. Multi-modal data is processed through two pipelines for text and image inputs. The resulting feature vector sequences are concatenated to provide the final prediction.

BERTBASE: For textual representations, the proposed approach uses the Bidirectional Encoder Representations from Transformers (BERT) technique. Text sequences from the datasets are fed to the BERTBASE model, which consists of twelve encoders. These encoders are individual transformer blocks with twelve self-attention heads. We use the model pre-trained on plain text from vast resources of English Wikipedia and BookCorpus. The choice of this model is made due to its advantage over word2vec and GloVe embeddings. BERT can easily differentiate among the contexts of various occurrences of the exact words, in which other embeddings are incapable. The input sequences from the datasets move sequentially through each of the twelve encoder layers. Self-attention is applied at each encoder layer, and the results are forwarded to the next encoder in sequence. The output is passed through a fully connected layer, and resulting data are the required textual feature vectors, F_t.

Multiscale CNN: Existing literature in fake news detection has primarily used pre-trained convolutional neural networks. This study proposes the use of multiscale CNNs for this task. This usage is supported by the fact that a single sequential convolutional network caters to working on specified input size. Whereas the advantage of multiscale CNNs lies in the allowance of running several convolutional models with different filter input sizes simultaneously. This lets the model analyze distinguished pixel regions. In practicality, the model can identify various objects in the image at a time with increased precision than the single-scale convolutional model. The image sequences from the datasets are fed to the proposed multiscale model consisting of two sequential convolutional networks. CNN 1 and CNN 2 consist of four and five 2D-convolutional layers, respectively. Each of these layers is

followed by two-dimensional max-pooling operational layers and ReLU activation function. Next, dense layers are added with shapes reducing consequently from 512 to 256 to 128 and then 2. To reduce overfitting, dropouts are used with both sequential CNNs of value 0.5. The outputs received from CNN1 and CNN2 are merged using concatenation, and fully connected layers are used thereafter, providing the final visual feature vectors, F_v.

Resulting feature vectors from textual feature extractor (F_t) and visual feature extractor (F_v) are combined using simple concatenation that produces multi-modal feature vectors. This output is the final classification result, F, that predicts the microblog into the real or fake category.

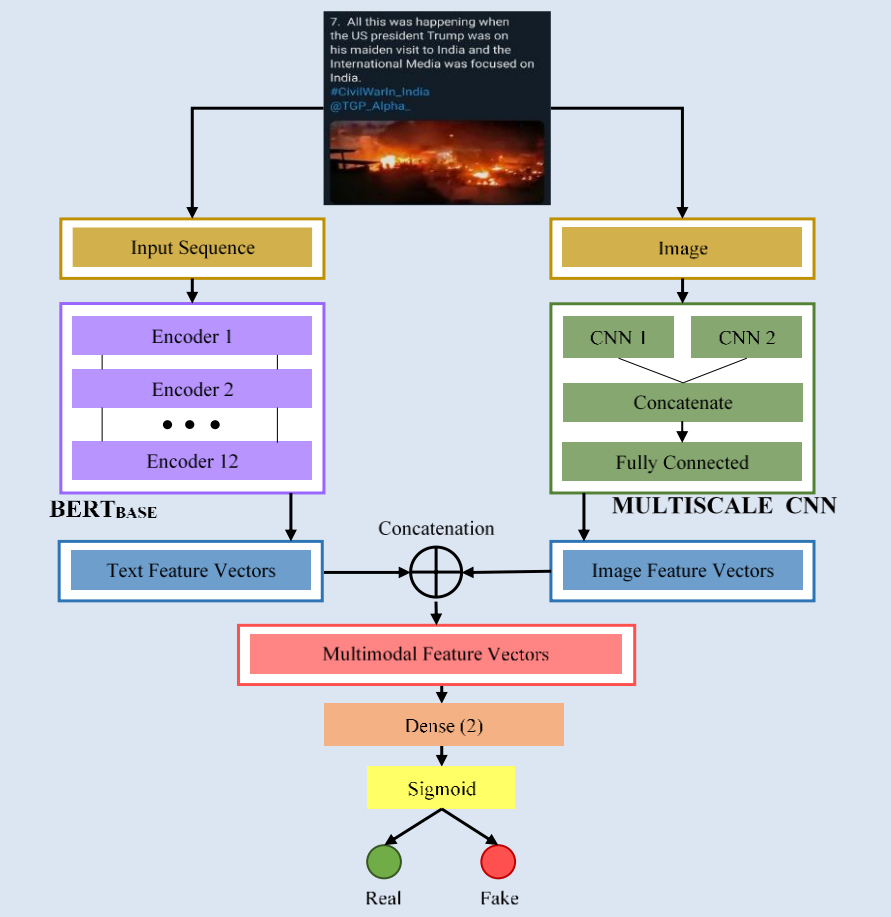


Figure 70: Proposed BERT-Multiscale CNN model for Fake News Detection

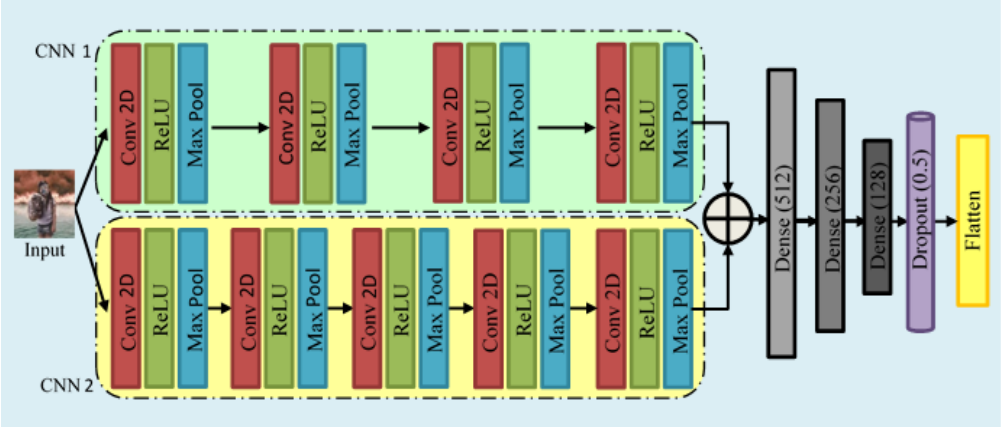


Figure 71: Architecture of Proposed Multiscale CNN

5.2 DATASETS AND PREPROCESSING

Experiments are performed on two standard publicly available real-world datasets: Twitter and Weibo. These datasets are collections of microblogs from social media networks. With the availability of multi-modal information in the microblogs, these datasets are appropriate for our work.

Twitter: This benchmark dataset was released as a part of the MediaEval 2016 [147] workshop for the 'Verifying Multimedia Use' task. The task aimed at detecting fake news on Twitter. The dataset consists of various information like text, image/video, and social context. The dataset consists of two sets: the development set and the testing set. We filter tweets containing both textual information and images. The development set is used for training and validation, and the algorithm is verified on the testing set.

Weibo: This dataset is a collection of microblogs from the authoritative news agency of China, Xinhua News Agency, and Weibo, a Chinese microblogging service. The dataset is a collection of microblogs collected between May 2012 to January 2016. The collection is verified by the official rumor debunking system of China. For the experiments, tweets along with their images are used. The dataset is split in a 4:1 ratio in accordance with previous literature and their investigations.

5.3 EXPERIMENTAL SETTINGS

Google Colab is used for performing all experiments using python 3. It allocated 12 GB NVIDIA TESLA K80 GPU and 13.53 GB of RAM. For NLP processing, the NLTK library

is used. Tokenization is performed using Regex. Stemming and Porter Stemmer and WordNet Lemmatizer aid lemmatization. For image classification, images are pre-processed to fixed input size of 224*224 to feed to 2D CNNs. The sigmoid function is used for supporting the classification into binary categories: real and fake. Adam optimizer is used for the visual model. Images are trained with a batch size of 64, the maximum which Google Colab could allocate in a runtime. The final result is calculated using four necessary metrics: accuracy, precision, recall, and f1-score.

5.4 EXPERIMENTAL RESULT ANALYSIS

In this section, we describe the achieved results on two datasets, Twitter and Weibo, and compare them with the results of existing state-of-the-art methods, Att-RNN, EANN, and TI-CNN, in terms of accuracy, precision, recall, and f1-score.

Table 23: Results and Baseline Comparison on Twitter and Weibo

Dataset	Method	Accuracy	Precision	Recall	F1-Score
Twitter	Att-RNN	0.664	0.749	0.615	0.676
	EANN	0.715	0.822	0.638	0.719
	TI-CNN	0.732	0.840	0.712	0.745
	Our Approach	**0.750**	**0.851**	**0.729**	**0.758**
Sina Weibo	Att-RNN	0.779	0.778	0.799	0.789
	EANN	0.827	0.847	0.812	0.829
	TI-CNN	0.831	0.839	0.826	0.844
	Our Approach	**0.914**	**0.870**	**0.891**	**0.878**

On the Twitter dataset, we achieve a real/fake classification accuracy of 75%. The precision, recall, and f1 scores are reported to be 85.1%, 72.9%, and 75.8%, respectively. For the Weibo dataset, the classification accuracy obtained is 91.4%. Precision, recall, and f1 scores are 87%, 89.1%, and 85.8%, respectively. As observable from the table, our approach surpasses the scores obtained by existing state-of-the-art techniques and provides better classification results. The approach, robust and quick to train, can perform efficient classification of fake and real news.

5.5 SUMMARY

This work proposes a novel approach for the veracity analysis of microblogs online. The proposed framework performs the classification of social media information into fake or

real. The framework components include a 12-encoder BERT for extracting textual features and a multiscale convolutional model with two CNNs for extracting visual features. The final prediction is made by concatenating the outputs obtained from BERT and Multiscale CNN model. The experimentation is performed on two publicly available real-world datasets, Twitter and Weibo. Results achieved via experimentation demonstrate the competence of the proposed framework. Our architecture has surpassed state-of-the-art methods in fake news detection. For future work, we aim to build more robust algorithms that could be useful for real-world applications to conquer the misinformation scenario.

CHAPTER 6

FAKE NEWS CHARACTERIZATION AND FACTOR IDENTIFICATION

6.1 INTRODUCTION

COVID-19 spread worldwide even faster than a human brain could imagine. Humans hardly even heard about it than before it turned to be the most fatal. After facing the catastrophic results only, many people became aware of it and started to ponder it. Talks about COVID-19 were everywhere and on everybody's minds and lips. Interactions about the hot topic have overwhelmed social networking platforms. Social media has now established its feet to feed information to people in the easiest way. The internet has been flooded with various types of information. But not everything that is on the internet is not reliable. Information that roams around on social media has not been validated and is merely people's ideas. Gradually these talks turned to be all sorts of fake news. With the feasibility of posting, sharing, and accessing the information on the web, its users can be quickly confounded with fake news. Fake news consists of every type of misinformation and disinformation. From the desks of politicians and public figures made the maiden attempt in spreading fake news worldwide, misleading people at large. It was the result of fake news that 5G towers in the UK turned into ruins. Fake news oozed out deadly political, social, religious, technological, environmental changes around the globe. It generated a sense of distrust among the people of the world. Enmity started grasping its enclosures. People claimed China to be the most causative element in spreading the disease. Detection of all sorts of talks that tend to be getting converted as fake news was the greatest need to lead the world into another mass destruction-like situation.

Fake news about the pandemic sprawled amongst various dimensions of society. One of these is the claiming of the remedial part of COVID-19. Enormous remedial approaches and suggestions started their part to play in contributing to fake news. "A pinch of turmeric or a drop of garlic juice could cure the fatal" was amongst the most prevailing unauthentic fake remedies. Poor perceptions, unproven methods, illogical claims, false figures, and alarming news overwhelmed the global information scenario. Social media platforms are well known for the spread of misinformation and denial of scientific literature [148]. False social media posts have also tricked users into relying on harmful and poisonous substances like weed, cannabis, and ethanol intake [149]. The rapid evolution of the COVID-19 pandemic has not

permitted immediate and specific scientific data [150]. COVID-19 is not the only fake news generating event. In the past, there have been many instances that led to colossal misinformation spread on online social networks, such as the 2016 US presidential elections, Pizzagate, hurricane Harvey, etc. [151]. COVID-19, whereas, is one major event generating misinformation on a scale larger than any other events. This led the World Health Organization into coining the term "Infodemic," referring to the mass propagation of false news revolving around the pandemic.

Previous research has contributed variously to solving the fake news problem. Researchers from behavioral sciences have covered the factors involved in sharing and accepting fake news [152, 153, 154]. Others have investigated several factors like user demographics and background information [155]. Many studies have developed fake news detection algorithms [156, 157]. Such algorithms widely utilize news content, such as linguistic features, visual features, and network features. However, there is an absence of ideal classifiers, and most of the fake news characteristics are unidentified. In this work, we identify several key factors associated with fake and real news on Twitter. We formulate fourteen hypotheses on the key elements and their direct and mediating relationship with fake news. These hypotheses are evaluated on two real-world datasets which contain tweets about the COVID-19 pandemic. MediaEval 2020 [158] is a benchmark dataset containing tweets pertaining to coronavirus and 5G conspiracy. CovidHeRA [159] is a collection of tweets associated with spreading health-related misinformation amidst the pandemic. The contribution of this work is the analysis of characteristics that differentiate between fake and real news. We identify the following key factors: sentiment polarity, gender, media usage, follower count, friends count, status count, retweet count, and favorites count. Interdependence of factors like sentiment polarity, gender, and media usage are studied intensely. The relationship between fake news and these factors has not been studied in past research. We also extend the work of Parikh et al. [160] by demonstrating the relationship between fake news and particular sentiment polarities. This work comes up with exciting outcomes suggesting important features demonstrating fake news dependence. The research bridges existing gaps in the literature and forms the basis for a new direction in fake news analysis. Our hypotheses shall be helpful in developing efficient fake news detection algorithms covering a wide range of fake news components.

6.2 RELATED WORKS

This section discusses the progress in fake news and hypothesis domain presented by fellow researchers so far.

Fake News: The menace of fake news has been a challenging problem for information consumers. It has constantly been a topic of concern in the research society. Various studies have discussed the identification and detection of fake news on online social networks [161]. Past studies have focused on a vast dimension of fake news ranging between its origin, propagation, consumption, and impact [162]. In the recent era, various solutions have been proposed to detect fake news by the help of exploiting its textual [163], visual [164], and nodal features [165]. In contrast, studies pertaining to hypothesis formulation and testing are very few. There is limited literature available discussing the latest trends in online social networks highlighting vulnerabilities in fake news propagation and consumption. It is essential to formulate and discover dependent dimensions of fake news. Some studies have proposed important insights beneficial for fake news detection. For instance, Parikh et al. proposed hypotheses discussing the origin, proliferation, and tone of fake information [166]. They concluded that such misleading information is published more on lesser-known websites than the popular ones. In terms of proliferation or sharing, unverified users are more often shared on social media than by verified accounts. They also demonstrated that fake news has a specific tone or sentiment (positive, negative, or neutral) but did not conclude which type of particular tone is fake news mostly related to. Their study provides ways to form additional hypotheses, which is also a motivation for our work. Demographics and culture form the basis of theories proposed by Rampersad and Althiyabi [167]. They identified the established relationships between age and acceptance of fake news. It was noted that other demographics like gender and education played a more minor role in fake news acceptance. Another notable hypothesis confirmed that educated people are less likely to accept fake news. It was also observed that culture indirectly impacts the acceptance of fake news significantly. Works have highlighted the connection between Third Person Effect (TPE) and fake news sharing [168, 169]. Brewer et al. have drawn several conclusions towards readers' reactions to consuming fake news [170]. Horne et al. have distinguished between real and fake news based on stylistic and physiological features of the text [171]. In another work by Silverman and SingerVine, it was identified that 75% of the US adults accepted fake news as true [172]. Similarly, Bovet and Makse studied the fake news propagation on Twitter during the 2016 US presidential elections and explored

its influence [173]. Altay et al. hypothesized the relation between users' reputation and fake news sharing [174]. They studied that very few people were indulged in sharing fake news and identified the causes of such behavior. They arrived at the conclusion that sharing fake news harmed people's reputations and resulted in trust issues, which is a significant reason for very few people being indulged in sharing fake news. Osatuyi and Hughes figured that the amount of information available on fake news platforms is lesser than real news [175]. Exploring the role of comments in identifying and rejecting fake news shows that users are less likely to accept fake news if they come across critical comments about the content [176].

Infodemic: With the outbreak of the COVID-19 pandemic, social media communication and interactions rose at a level greater than before. Global concerns about the disease brought the world together to share information on online social networks. Such large-scale propagation gave rise to a phenomenon- "Infodemic." In an early response, researchers approached this problem by analyzing various concerns and suggesting solutions to the issue. Moscadelli et al. [177] have investigated the topics about the pandemic most polluted with fake news. Calvillo et al. [178] have analyzed political associations with the discerning of fake news. Hypotheses linking the fake news belief structure to its acceptance, Kim and Kim [179] proposed that factors like source credibility, quality of information, receiver's ability, perceived benefit, trust, and knowledge decrease people's belief in fake news. Contrastingly, heuristic information, perceived risk, and stigma strengthen the confidence in fake news. Greene and Murphy [180] have discussed the likeliness of people sharing true or false stories on social media, establishing the association with their knowledge concerns. Another study that links conscience and ideology with infodemic sharing behavior is provided by Lawson and Kakkar [181]. Montesi [182] spreads light on the nature of infodemic and suggests that the harm caused by fake news is not health-related but more of a moral sort. Society, politics, and society are identified as the dominant infodemic themes. Building constructs over the Third Person Effect (TPE), Lui and Huang [183] have facts regarding the susceptibility and perception of fake news in the pandemic era. Similarly, Laato et al. [184] discuss the factors such as information sharing, information overload, and cyberchondria aiding fake news propagation. Experimenting on a Nigerian sample, Sulaiman [185] proposed no relationship between information evaluation and fake news sharing. With many hypotheses, Alvi and Saraswat [186] explored connections amongst various heuristic and systematic factors such as Sharing Motivation, Social Media Fatigue, Feel Good Factor. Fear Of Missing Out. News Characteristics, Extraversion, Conscientiousness, Agreeableness, Neuroticism, Trust, and

Openness. As observed from the existing literature, past studies revolve around identifying psychological and behavioral factors that demonstrate any relationship with fake news. There is a research gap in characterizing features that could aid in distinguishing false information from real and serve as contributing factors to build fake news detection algorithms.

6.3 RESEARCH METHODOLOGY

Data: This study uses two publicly available benchmark datasets, MediaEval 2020 [187] and CovidHeRA [188]. MediaEval 2020 issued a benchmark dataset for its fake news detection task. The dataset consists of 5842 tweets classified into three classes: 5G coronavirus conspiracy, other conspiracy, and non-conspiracy. The tweets contain real and false information revolving around the COVID-19 pandemic. For this study, we classify these tweets into two coarse classes, with non-conspiracy tweets as real and the remaining tweets as fake. CovidHeRA is another benchmark dataset containing false tweets related to coronavirus and health. These tweets are a collection of fake remedies, preventive measures, treatments, and other health-related information spread across Twitter amidst the pandemic. Originally, the datasets consisted of tweet ids.

Table 24: Count of fake and real items with gender as a category

	CovidHeRA			Mediaeval		
Label	Male	Female	Total	Male	Female	Total
Fake	1532	772	2304	929	837	1766
Real	42683	40104	82787	2011	2065	4076
Total	44215	40876	85091	2940	2902	5842

Table 25: Count of fake and real items with sentiment polarity as a category

	CovidHeRA				Mediaeval			
Label	Negative	Neutral	Positive	Total	Negative	Neutral	Positive	Total
Fake	1292	391	621	2304	1042	346	378	1766
Real	31004	24638	27145	82787	2320	690	1066	4076
Total	32296	25029	27766	85091	3362	1036	1444	5842

Table 26: Count of fake and real items with media usage as a category

	CovidHeRA			Mediaeval		
Label	With	W/o Media	Total	With	W/o Media	Total
Fake	150	2154	2304	289	1477	1766
Real	17700	65087	82787	791	3285	4076
Total	17850	67241	85091	1080	4762	5842

To procure various characteristics of the tweets, the python library Tweepy is utilized. This scraping results in providing various information of the tweet and user content. This contextual information forms the basis of this study. To obtain the gender information of Twitter users, a gender predictor algorithm by Sap et al. [189] is used. Sentiments on the dataset are extracted using Microsoft's Text Analytics service. Sentiment scores are returned as values in the range of 0.0 to 0.1. A score between 0.0 to 0.3 signifies negative, 0.3 to 0.7 represents neutral and 0.7 to 1.0 represents positive sentiment. For media usage, we utilize the 'extended_entities' column from the scraped datasets. Sizes of both the datasets pertaining to each category are provided in tables 14, 15, and 16.

Research Hypothesis: To identify characteristics that distinguish fake news and real news and consequently identify fake news based on these characteristics, we have formed fourteen hypotheses based on the qualitative and quantitative variables present in the dataset. Past research to determine factors related to fake news is limited. To identify the dependence of social media misinformation, we identify and analyze eight key elements: sentiment polarity, gender, media usage, follower count, friends count, status count, retweet count, and favorite count. We assume that fake news characterization, propagation, and acceptance have a relationship with these factors, which can be consequently utilized in fake news detection. For a better understanding, each tweet labeled as fake/real in the datasets has specific characteristics mentioned above. It is crucial to examine which feasible aspects demonstrate a relationship with false tweets. We also aim to study if there are any significantly different factors between real and fake tweets. By establishing such relationships, we tend to describe certain features useful for real and fake tweet classification. As evident from the existing literature, very few features have been exploited by fake news detection algorithms. Now examining the stated features, we propose to add more of such contributing characteristics. Qualitative hypotheses $H_{A,}$ $H_{B,}$ and $H_{C,}$ are tested to scrutinize the direct relationships between sentiment, gender, and media usage with fake news, respectively. Further, it is vital to analyze if the bias of one independent variable influences the bias of another independent variable. For example, to test whether or not it is the higher proportion of one categorical variable contributing to the higher proportion of another categorical variable. To do so, we construct six more qualitative hypotheses, H_D, H_E, H_F, $H_{G,}$ H_H, and H_I. These nine hypotheses are tested using the Chi-square test of independence. To study quantitative variables, we formulate hypotheses H_J to H_N and perform Analysis of Means on each one of, also and calculate intervals.

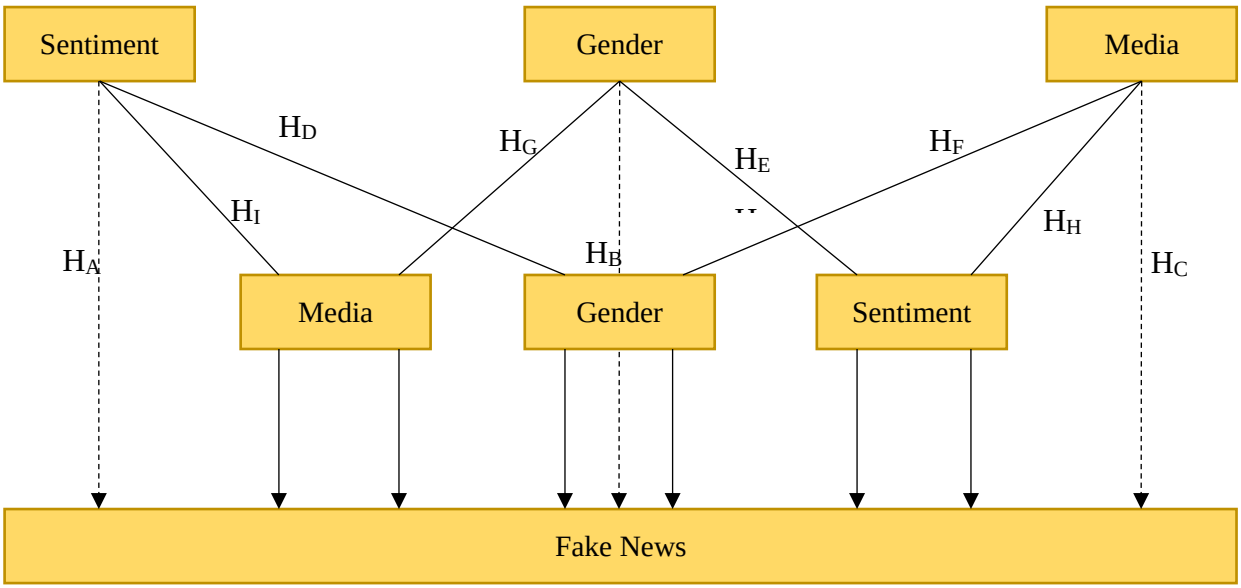


Figure 72: Factors determining fake news (qualitative hypotheses)

Qualitative Hypotheses and Factors

Sentiment: According to Parikh et al., it is widely assumed that most of the news spreading online is negative in terms of its linguistic tone. However, it has not been proven that fake news has a higher negative polarity than neutral or positive polarities. Parikh et al. noted that it was inconclusive to say if fake news had a bias towards a particular polarity. Following their assumption, H_A forms the primary hypothesis to test if fake news has a tendency towards a specific sentiment polarity.

H_{A0}: There is no bias in the proportion of different sentiments between fake news and real news.

H_{A1}: There is a significant bias in the proportion of different sentiments between fake news and real news.

Gender: Rampersad and Althiyabi, examining a sample of Saudi Arabia, observed that gender has a weakly positive effect on the acceptance of fake news by people. The sample is specific to a particular demographic region. In our research, datasets consist of tweets from Twitter users across the globe. This helps to examine the assumptions on a universal scale. We test this hypothesis by using HB's statement to verify if there is a significant relationship between gender and false information.

H_{B0}: There is no bias of the gender of users involved in fake news with respect to real news.

H_{B1}: There is a significant bias of gender of users involved in fake news with respect to real news.

Media: Several fake news detection algorithms have been designed that detect whether a visual media in a piece of fake information is credible or not. We, hereby, analyze whether it can be stated solely based on the presence of visual media that a post/message is false. We categorize the datasets into two modalities: without and with visual media (pictures/videos). We try to analyze what data modality of social media posts contribute more/demonstrate bias towards misinformation using the statement H_C.

H_{C0}: There is no bias of media usage in fake news with respect to real news.
H_{C1}: There is a significant bias of media usage in fake news with respect to real news.

Based on the above three univariate hypotheses, we decide the mediating relationships among these factors and formulate multivariate hypotheses (H_D to H_I) to determine whether bias in one of the above proportions is due to bias in proportions of the other variable.

H_{D0}: There is no influence of bias in the proportion of a particular gender of the user on the bias in the proportion of sentiments in fake news with respect to real news.
H_{D1}: There is significant influence of bias in the proportion of a particular gender of the user on the bias in the proportion of sentiments in fake news with respect to real news.

H_{E0}: There is no bias in the proportion of a particular sentiment used in fake news between different gender of users.
H_{E1}: There is a significant bias in the proportion of a particular sentiment used in fake news between different gender of users.

H_{F0}: There is no bias in inducing a particular sentiment with media usage in fake news.
H_{F1}: There is a significant bias in inducing a particular sentiment with media usage in fake news.

H_{G0}: There is no bias in the usage of media amongst different sentiments used in fake news.
H_{G1}: There is a significant bias in media usage amongst different sentiments used in fake news.

H_{H0}: There is no relationship between a particular gender and media usage in fake news.

H_{H1}: There is a significant relationship between a particular gender and media usage in fake news.

H_{I0}: There is no bias in and media usage in fake news between different gender of users.
H_{I1}: There is a significant bias in and usage of media in fake news between different gender of users.

Quantitative Hypotheses and Factors

Using the data scraped from Twitter, we decided on testing our hypotheses on five key factors, which can be categorized into three user/profile-specific features, i.e., the number of followers, friends, and statuses and two post-specific features, i.e., retweets count and favorites count. In our novel approach, we assume that these factors can be utilized in identifying the credibility of tweets, or in other words, labeling of tweets. Moreover, we assume that these factors impose an effect on fake news sharing and acceptance.

Followers and friends count determine the extent of reachability of a particular post or message within the user's social network who created it. A retweet is an action of sharing a particular tweet on one's timeline, which is done mainly by the follower of the user who created it and is visible to other Twitter users who turn the followers of the user who retweeted it. Retweet count determines the propagation and acceptance behavior of a fake post by checking the social reach. It is similar to the action "Share" on other social networks. It spreads a particular post to the user's social network. The larger the retweet count, the more likely the people reading the post will believe that particular piece of information and further spread it across the web. Status count corresponds to the number of total posts/retweets a specific user has posted since the creation of his account. Favorites are user markings made on a post a user would like to save for the future.

We determine the relationship between these quantitative variables and the label of the post, i.e., the relationship between the number of retweets and favorites of the post and the followers, friends, and status of the user who posted it, and it being real or fake. Since the source of misinformation can range from a random regular user to a credible account such as commercial news channels, journalists, or celebrities, it becomes difficult to assume any specific range for the count of these quantitative variables. Hence, we test based on a characteristic whether there is a significantly distinguishable bias in the values attributed to the mean and a confidence interval around it for each of these variables. In other words, the

probability with which a post or a piece of information under examination can be labeled as fake or real based on its values of the above-mentioned quantitative variables.

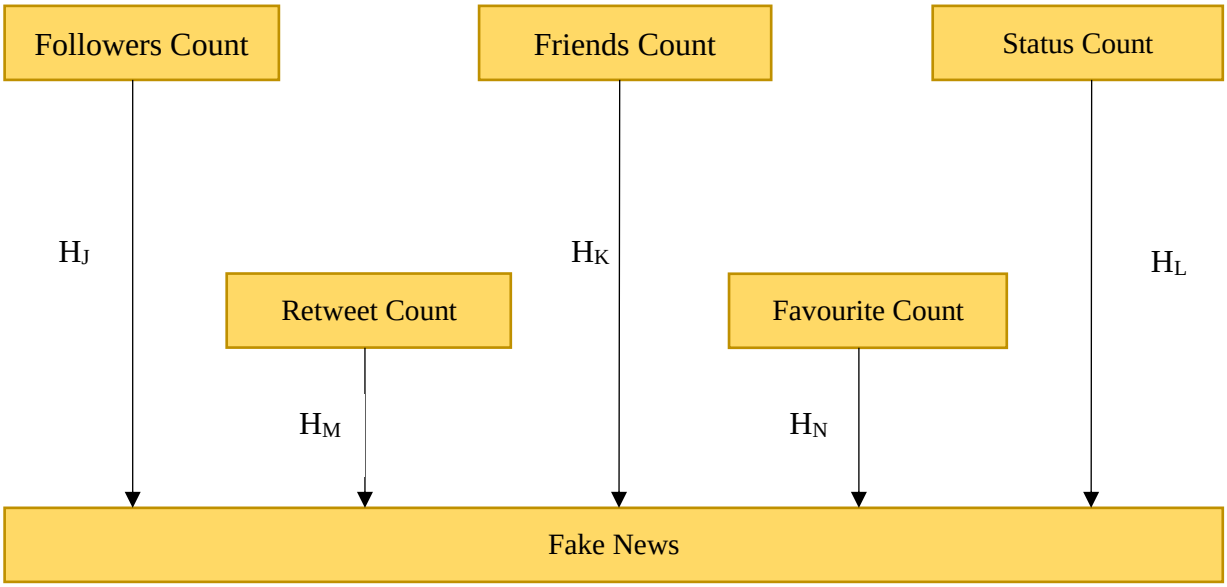


Figure 73: Factors determining fake news (quantitative hypotheses)

H_{J0}: There is no bias of follower count in fake news.

H_{J1}: There is a significantly distinguishable bias of follower count in fake news.

H_{K0}: There is no bias of friends count in fake news.

H_{K1}: There is a significantly distinguishable bias of friends count in fake news.

H_{L0}: There is no bias of status count in fake news.

H_{L1}: There is a significantly distinguishable bias of status count in fake news.

H_{M0}: There is no bias of retweet count in fake news.

H_{M1}: There is a significantly distinguishable bias of retweet count in fake news.

H_{N0}: There is no bias of favorite count in fake news.

H_{N1}: There is a significantly distinguishable bias of favorite count in fake news.

6.4 RESULTS

To test on the nine hypotheses H_A to H_I, which are formed upon the categorical variables, we use the Chi-Square test of independence alongside computing "Cramer's V," "Pearson's r," and "spearman's rho" values. Cramer's V value provides us with the strength of association between the nominal categorical variables for the conclusion arrived using the Chi-Square test. Its values range between 0 and 1. Pearson's r value signifies both the strength of

association and the direction of the association between two continuous variables. Here direction indicates if one variable would increase or decrease with respect to change in another variable. Its values range from -1 to +1, where the value of -1 means that as one variable increases, the other decreases, and +1 means that as one variable increases, the other increases too. A value of 0 indicates no strength of association. Spearman's rho values differ from the outcomes of Pearson's r values by a feature that they can describe the correlation even when the variables do not have a linear association. It is also proof from the long tail of outlier values as it uses the ranks of the values of the variable. The values in the table 17 include degrees of freedom as df, Chi-Square test value as χ^2, probability value as p-value and Cremer's V value, Pearson's r-value, and Spearman's rho. The first column in this table indicates the hypothesis to which the variables and their values belong to. From the first row of the same table, we observe that χ^2 values for testing hypothesis H_A with 2 degrees of freedom (df) for both CovidHeRA and MediaEval datasets are 352.963 and 17.103, respectively, and are more significant than critical value $\chi^2_c = 5.991$ with $p < 0.001$ (Significance level $\alpha = 0.05 = p_c$, critical p-value). This implies that there is a significant difference in proportions of sentiments used between Fake and Real news. But despite there being a substantial difference in ratios, low values of Cramer's V (less than 0.2), Pearson's r (between -0.20 and +0.20), and Spearman's rho (between -0.20 and +0.20) indicate weak association of label (news being fake or real) and the sentiment (sentiment being negative or neutral or positive). These values (Cramer's V, Person's r, and Spearman's rho) are low for all the hypotheses tested. Therefore, we rely on comparing Actual values with Expected values to determine the association between an independent and a categorical dependent variable, or in other words, the bias of fake news towards a specific or a group of categorical variables. We observe that in both CovidHeRA and MediaEval datasets, Fake news with Negative sentiment has a higher Actual count (1292, 1042) with respect to Expected count (874.5, 1016.3) and Fake news with Positive sentiment has less Actual count (621, 378) with respect to Expected count (751.8, 436.5). Count of Neutral sentiment varies inversely in both datasets, with CovidHeRA showing reduced count and MediaEval showing an Increase. Similarly, we observe from the same tables that the Actual count of Real news with Negative sentiment is less than that of the Expected count in both datasets. The Actual count of Real news with Positive sentiment is greater than that of the Expected count in both the datasets. Therefore, we reject the Null hypothesis (H_0) of H_A and observe that Fake news propagation during CoVID-19 has had a proportional bias towards Negative sentiment.

Table 27: Expected count of fake and real items with gender as a category

	CovidHeRA			Mediaeval		
Label	Male	Female	Total	Male	Female	Total
Fake	1197	1107	2304	888.7	877.3	1766
Real	43018	39769	82787	2051.3	2024.7	4076
Total	44215	40876	85091	2940	2902	5842

Table 28: Expected count of fake and real items with sentiment polarity as a category

	CovidHeRA				Mediaeval			
Label	Negative	Neutral	Positive	Total	Negative	Neutral	Positive	Total
Fake	874.5	677.7	751.8	2304	1016.3	313.2	436.5	1766
Real	31421.5	24351.3	27014.2	82787	2345.7	722.8	1007.5	4076
Total	32296	25029	27766	85091	3362	1036	1444	5842

Table 29: Expected count of fake and real items with media usage as a category

	CovidHeRA			Mediaeval		
Label	With	W/o Media	Total	With	W/o Media	Total
Fake	483	1821	2304	326.5	1439.5	1766
Real	17367	65420	82787	753.5	3322.5	4076
Total	17850	67241	85091	1080	4762	5842

We observe that χ^2 values for testing hypothesis H_B for both CovidHeRA and MediaEval datasets are 200.321 and 5.261, respectively, and are greater than critical value $\chi^2_c = 3.841$ with $p < 0.001$ and $p = 0.022$, respectively, both less than $\alpha = 0.05$. This implies a significant difference in proportions of the gender of users between Fake and Real news. By comparing Actual values with Expected values, we observe that the Male gender has a greater Actual proportion in Fake news than the Expected proportion, and the Female gender has a higher Actual proportion involved in Real news than Expected Proportion, in both datasets. Therefore, we reject the Null hypothesis (H_0) for H_B and observe a significant bias in the gender of users involved in CoVID-19 Fake news propagation.

To test for Hypothesis H_C, we observe that χ^2 values for both CovidHeRA and MediaEval datasets are 298.995 and 7.765, respectively, and are greater than critical value $\chi^2_c = 3.841$ with $p < 0.001$ and $p = 0.006$ respectively, both less than $\alpha = 0.05$. For both the datasets, comparing the values of Actual and Expected Media usage shows that Actual values for Fake news with media used is less than Expected values and the same is more in the case of Real news. Therefore, there is a significant difference in the proportion of Fake news and Real news propagation with media usage than the expected proportion, which leads us to reject the Null Hypothesis (H_0) for H_C.

The test for hypothesis H_D, the χ^2 values of Male gender from datasets CovidHeRA and MediaEval are 217.67 and 13.342, respectively, both higher than $\chi^2_c = 5.991$ and p values being $p < 0.001$ and $p = 0.001$, respectively, both less than $\alpha = 0.05$. From row 5, the χ^2 value for Female gender from CovidHeRA dataset is 169.979, greater than the critical value $\chi^2_c = 5.991$ and the value of $p < 0.001$ is less than $\alpha = 0.05$. But for the same gender in the MediaEval dataset, the χ^2 value turns out to be 5.503, which is less than $\chi^2_c = 5.991$, and the p-value of $p = 0.064 > \alpha = 0.05$ suggests contradictory inference from these two datasets. But since the MediaEval dataset gave both the χ^2 and p values close to their respective critical values for female gender, we reject Null Hypothesis (H_0) for H_D and conclude that there is a significant bias in proportion of sentiments used by users of both the gender and the bias in proportion of the user gender has no influence on the bias of proportion of sentiments.

Further, to identify towards which sentiment is the bias more by the users of both genders, we use results from rows six, seven, and eight for testing hypothesis H_E. For the CovidHeRA dataset, the three rows mentioned above have χ^2 value of 78.005, 13.65, and 146.509 for negative, neutral, and positive sentiment, respectively, which are all greater than $\chi^2_c = 3.841$ and their respective p values being $p < 0.001$ for all three, is less than $\alpha = 0.05$. Results from this dataset do not indicate the specific sentiment towards which the bias is more. However, we can infer that there is a significant difference in the proportion of each sentiment when compared to real news. Observing results from these three rows for the MediaEval dataset, we obtain χ^2 value of 1.702, 4.82, and 0.411, for negative, neutral, and positive sentiments, respectively, where χ^2 values for Negative and Positive sentiments are both less than $\chi^2_c = 3.841$ and for Neutral sentiment, the χ^2 value is higher than χ^2_c. The p values for these corresponding χ^2 values are $p = 0.192$, $p = 0.028$ and $p = 0.521$, respectively. This shows no significant bias of the user gender on Negative and Positive sentiment as p values (0.192 and 0.521) obtained are greater than $\alpha = 0.05$. But for Neutral sentiment, we observe a bias as the p-value of 0.028 is less than $\alpha = 0.05$. Therefore, we reject the Null hypothesis (H_0) for H_E and conclude that Fake news is more biased towards being sentiment Neutral, followed by being sentiment Negative, and show no significant difference in proportions of Real news towards being sentiment Positive.

For testing Hypothesis, H_F, the bias of usage of media to induce a particular sentiment in the propagation of COVID-19 Fake news, the values from rows nine, ten and eleven for CovidHeRA dataset indicate χ^2 values of 97.382, 59.615, and 124.61 for Negative, Neutral and Positive sentiment, respectively, with all of them being greater than $\chi^2_c = 5.991$ and with a p-

value for each of them being $p < 0.001$, less than $\alpha = 0.05$ indicate rejection of Null Hypothesis (H_0) for H_F. For the MediaEval dataset, however, the χ^2 values of 0.235, 2.399, and 20.077 for Negative, Neutral and Positive sentiments, respectively, with the former two being less than $\chi^2_c = 3.841$ and the latter being more excellent, and their respective p values being $p = 0.628$, $p = 0.121$ and $p < 0.001$ indicate that only for Positive sentiment, there is a significant difference of proportion in the usage of media for Fake news with respect to real news. From the contradictory results from the two datasets for Negative and Neutral sentiments, we understand that there is a bias produced by usage of media for only positive sentiment. Hence, we reject the Null Hypothesis (H_0) for H_F.

Table 30: Chi-square test on qualitative hypotheses

	Datasets →	CovidHeRA						MediaEval					
(H)	variables ↓	df	χ^2	p value	V^2	r	rho	df	χ^2	p value	V^2	r	rho
H_A	Label vs Sentiment	2	352.963	p < 0.001	0.004	0.047	0.048	2	17.10	p < 0.001	0.003	0.037	0.032
H_B	Label vs Gender	1	200.321	p < 0.001	0.002	0.048	0.048	1	5.261	p = 0.022	0.001	0.03	0.03
H_C	Label vs Media usage	1	298.995	p < 0.001	0.003	0.059	0.059	1	7.565	p = 0.006	0.001	0.035	0.035
H_D	Label vs Sentiment (Gender - male)	2	217.67	p < 0.001	0.004	0.039	0.041	2	13.342	p = 0.001	0.004	0.035	0.027
H_D	Label vs Sentiment (Gender - female)	2	169.979	p < 0.001	0.004	0.058	0.058	2	5.503	p = 0.064	0.001	0.038	0.035
H_E	Label vs Gender (Sentiment - Negative)	1	78.005	p < 0.001	0.002	0.049	0.049	1	1.702	p = 0.192	0.001	0.023	0.023
H_E	Label vs Gender (Sentiment - Neutral)	1	13.65	p < 0.001	0.001	0.023	0.023	1	4.82	p = 0.028	0.004	0.068	0.068
H_E	Label vs Gender (Sentiment - Positive)	1	146.509	p < 0.001	0.005	0.072	0.072	1	0.411	p = 0.521	0	0.016	0.016
H_F	Label vs Media usage (Sentiment - Negative)	1	97.382	p < 0.001	0.003	0.055	0.055	1	0.235	p = 0.628	0	-0.008	-0.008
H_F	Label vs Media usage (Sentiment - Neutral)	1	59.615	p < 0.001	0.002	0.048	0.048	1	2.399	p = 0.121	0.002	0.048	0.048
H_F	Label vs Media usage (Sentiment - Positive)	1	124.61	p < 0.001	0.004	0.066	0.066	1	20.077	p < 0.001	0.013	0.117	0.117
H_G	Label vs Sentiment (Media not used)	2	267.346	p < 0.001	0.003	0.042	0.044	2	7.223	p = 0.027	0.001	0.01	0.005
H_G	Label vs Sentiment (Media used)	2	61.585	p < 0.001	0.003	0.045	0.042	2	25.154	p < 0.001	0.023	0.145	0.141
H_H	Label vs Gender (Media not used)	1	193.333	p < 0.001	0.003	0.053	0.053	1	5.561	p = 0.018	0.001	0.034	0.034
H_H	Label vs Gender (media used)	1	5.472	p = 0.019	0	0.017	0.017	1	0.345	p = 0.557	0	0.017	0.017
H_I	Label vs Media usage (Gender - male)	1	209.649	p < 0.001	0.005	0.069	0.069	1	5.664	p = 0.017	0.001	0.043	0.043
H_I	Label vs Media usage (Gender - female)	1	87.831	p < 0.001	0.002	0.046	0.046	1	2.529	p = 0.112	0.001	0.03	0.03

Statistics	Followers		Friends		Retweets		Status		Favorites	
	Fake	**Real**	**Fake**	**Real**	**Fake**	**Real**	**Fake**	**Real**	**Fake**	**Real**
Mean (C)	5421.657	63656.21	3181.374	2293.652	154.132	628.718	56262.98	46189.4	2.238	7.766
Standard Error (C)	445.735	3847.024	149.150	35.658	38.207	17.616	2269.618	497.010	0.255	0.490
Median (C)	1742.5	21733	953	605	52	173	17180	9238	1	2
Mode (C)	706	7810	775	209	18	28	3145	1760	0	0
Standard Deviation (C)	21395.32	1106894	7159.233	10259.94	1833.94	5068.56	108941.7	143003.4	12.261	141.126
Sample Variance (C)	4.58E+08	1.23E+12	51254619	1.05E+08	3.36E+06	2.57E+07	1.19E+10	2.04E+10	150.344	19916.57
Count (C)	2304	82787	2304	82787	2304	82787	2304	82787	2304	82787
Confidence Level(95.0%) (C)	874.085	7540.138	292.483	69.890	74.886	34.527	4450.709	974.136	0.500	0.961
Mean (M)	23255.37	99511.34	3012.989	1999.394	260.701	644.781	38369.96	55846.16	679.669	2244.092
Standard Error (M)	9979.619	11302.57	302.646	159.182	71.410	61.498	1787.307	1744.372	201.229	224.780
Median (M)	4711.5	37160.5	733	609	60	155	12955	18305.5	48	292
Mode (M)	1180	12548	650	138	77	92	547	446	48	177
Standard Deviation (M)	419381.5	721596.4	12718.35	10162.77	3000.955	3926.28	75109.43	111366.9	8456.42	14350.77
Sample Variance (M)	1.76E+11	5.21E+11	1.62E+08	1.03E+08	9005729	15415676	5.64E+09	1.24E+10	71511039	2.06E+08
Count (M)	1766	4076	1766	4076	1766	4076	1766	4076	1766	4076
Confidence Level(95.0%) (M)	19560.05	22153.04	593.583	311.998	140.058	120.570	3505.461	3419.921	394.672	440.692

Table 31: Descriptive statistics of CovidHeRA(C) and MediaEval dataset(M)

We test for Hypothesis H_G to observe a bias of proportion of sentiment caused when media is used and when it is not used, respectively. For CovidHeRA dataset, for with usage of media (row 12) and without the usage of media row (13), χ^2 values of 267.346 and 61.585, respectively, both less than $\chi^2_c = 5.991$ and their respective p values of $p < 0.001$ each for both being less than $\alpha = 0.05$, suggest that there is a difference in the proportion of sentiment used in Fake news with respect to Real news. Similar inference can be obtained from MediaEval dataset, in which, with the usage of media (row 12) and without the use of media row (13) have χ^2 values of 7.223 and 25.15, respectively, both less than $\chi^2_c = 5.991$ and their respective p values of $p = 0.027$ and $p < 0.001$, both being less than $\alpha = 0.05$. Hence, there is a bias induced in the proportions of sentiment in Fake news with respect to Real news by usage and non-usage of media, and therefore we reject the Null Hypothesis (H_0) for H_G.

We test for Hypothesis H_H to check for bias in proportion of gender of Fake news with respect to Real news is influenced by bias in usage of media. For CovidHeRA, we obtain χ^2 values of 193.333 and 5.472 for "media used" and "media not used", respectively, both greater than $\chi^2_c = 3.84$ with their respective p values being $p < 0.001$ and $p = 0.019$, both less than $\alpha = 0.05$. For the MediaEval dataset, for the same rows, we obtain χ^2 values of 5.561 and 0.345 and p values of $p = 0.018$ and $p = 0.557$ for "media used" and "media not used," respectively. We observe that for "media not used," the test shows the opposite result with that compared from CovidHeRA dataset, meaning that there is no difference in the proportion of user's gender when media is not used in Fake news propagation, with respect to Real news propagation. These contradictory results make Hypothesis H_H inconclusive.

For the CovidHeRA dataset, both genders show that there is a difference in the proportion of media used for Fake news propagation with respect to Real news. This can be observed as the χ^2 values of 209.649 and 87.831 for the male and female gender, respectively, are both greater than $\chi^2_c = 3.84$, and their respective p values, both $p < 0.001$ is more diminutive than $\alpha = 0.05$. In the MediaEval dataset, we observe that while users of Male gender with χ^2 value of 5.664 and $p = 0.017$ show difference in the proportion of media used for Fake news with respect to Real news, but for Female gender, indifference in proportions of usage of media in Fake news with respect to real news is observed as the χ^2 value of 2.529 is less than $\chi^2_c = 3.84$ and its p-value of $p = 0.112$ is more remarkable than $\alpha = 0.05$. Therefore, for Hypothesis H_I, we cannot come to any conclusive decision.

For the Quantitative variables, we plot the data distribution around the mean with a 95% confidence interval. This will distinguish the central values of the variables and help us determine the strength of the distinguishment, i.e., the smaller the upper and lower bound distance from the mean, the more the reliance on these values representing the true mean value of the population. We observe that users who propagated Fake news have a smaller number of followers than the users with Real news. The mean values for Fake news for these datasets are 5421.65 and 23255.37 in the order mentioned above. These are distinct from the mean number of followers of Real news 63656.21 and 99511.34 for the two datasets. We also observe that there is a significant bias in the number of followers of the users of Fake news and Real news as the range of 95% CI for mean do not overlap for Fake news and Real news and therefore attributing a label to a piece of information on Twitter by comparing, the number of followers of the user who shared it with the mean range of these plots can be done more accurately.

From the plots of the number of Friends in fig 3 and fig 4 for CovidHeRA and MediaEval datasets, respectively, the previously mentioned inference becomes much more robust as not only the 95% CI bounds remain distinct for Fake news and real news, but also the closer proximity of the value of mean for a particular label in both datasets shows the repeatability of the trend. The mean value for Fake news in CovidHeRA and MediaEval dataset is 3181.374 and 3012.989, respectively, and the same for Real news in these datasets is 2293.652 and 1999.394, respectively. There is a significant bias in the mean number of friends for users who propagated Fake news compared to the number of friends of users who propagated real news.

The plots from fig 4 and fig 4 for the number of retweets have similar mean value for fake news and real news. For Fake news, the mean values of 154.132 and 260.701 for the two datasets, and Real news, the mean values are 628.718, and 644.781 show the closeness within the label and distinction between the labels. Therefore, this bias can prove helpful to label a piece of information based on its proximity to one of the mean values 95% CI interval.

For the “number of statuses” variable, the 95% CI interval for mean and the mean value for Fake news and Real news alternate between the two datasets. Therefore, we cannot come to any specific conclusion using the information of this variable of a particular information sample despite there being bias in the mean values between the Labels. The same conclusion can be drawn for “number of favorites” as the ranges in both the datasets are significantly

different amongst the same variable. Hence, any information about this variable in a sample information under test cannot be determined as Fake or Real.

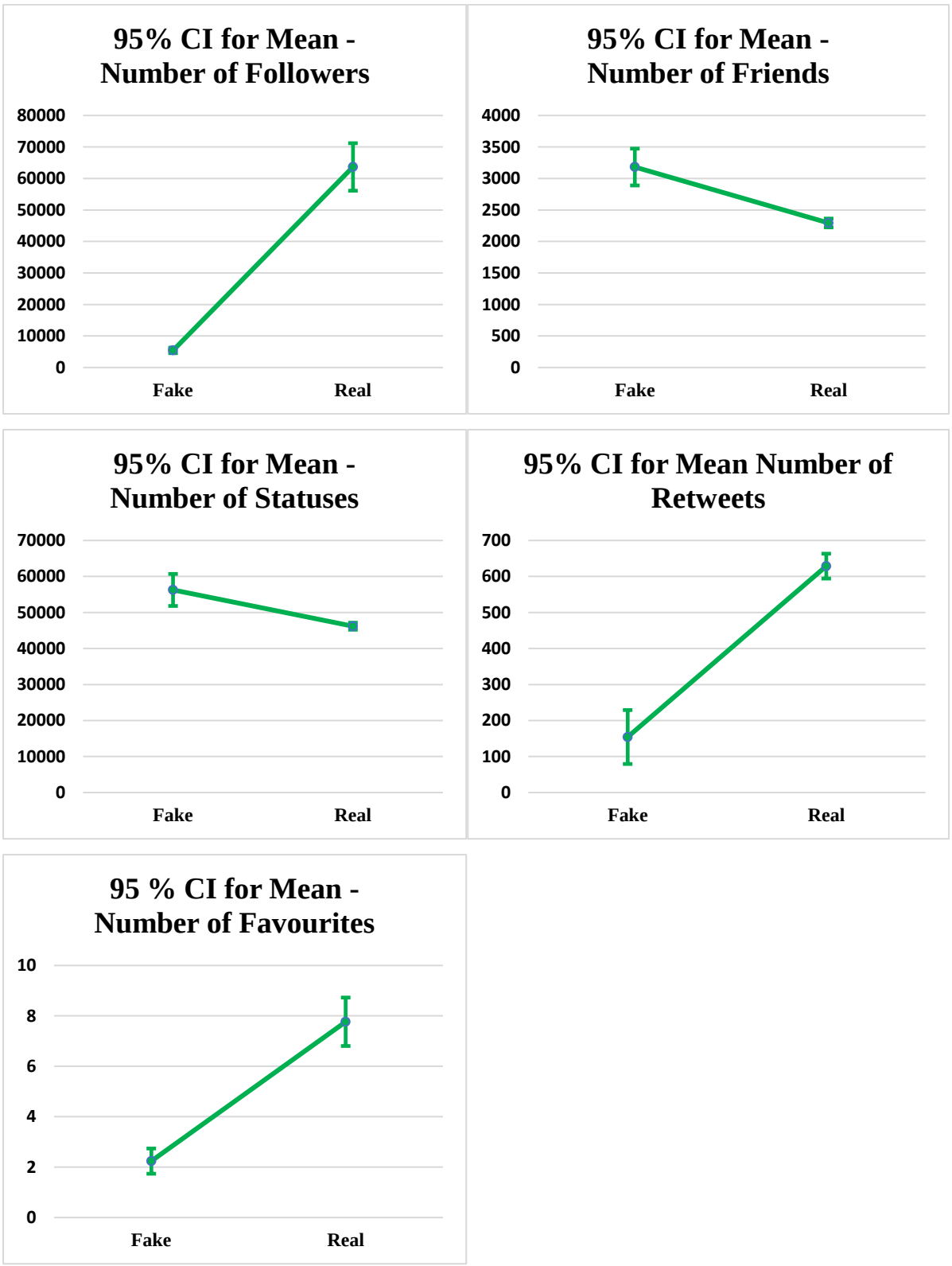


Figure 74: 95% Confidence Interval for quantitative factors on CovidHeRA dataset

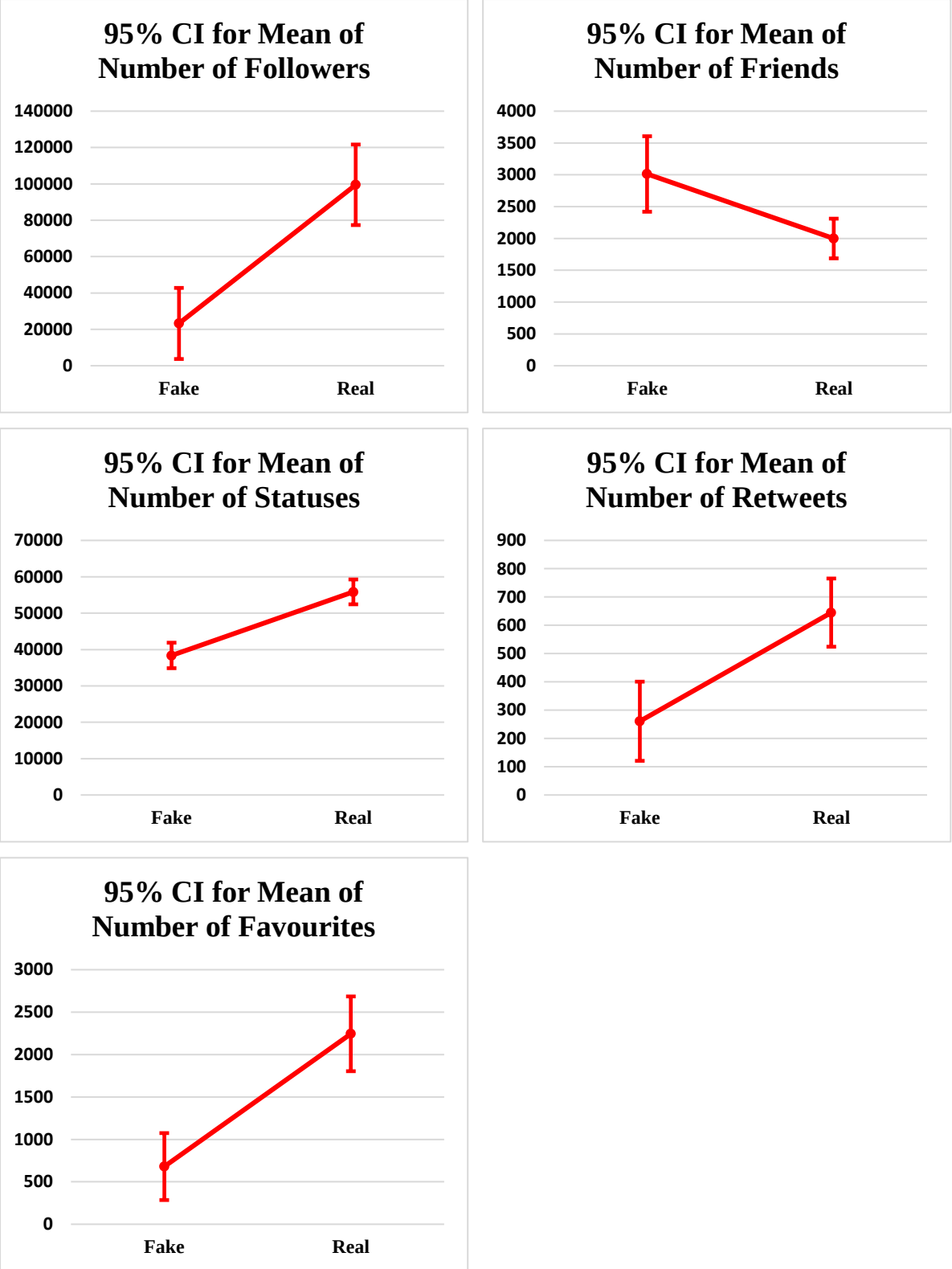


Figure 75: 95% Confidence Interval for quantitative factors on MediaEval dataset

6.5 DISCUSSION

Fake news on social media is a menace hard to identify and characterize. It is unclear which factors are helpful in distinguishing between real and fake news. Past literature has identified several psychological and behavioral features associated with fake news propagation and acceptance. Little research has been done in identifying key factors characterizing fake news. This study delves deep into factor analysis and their interdependence. We examine how certain factors influence fake news detection and propagation on Twitter. Table 22 summarizes the results of all hypotheses considered.

Table 32: Summary Table

Hypotheses	Results
H_A: Bias of sentiment in fake news with respect to real news.	Reject Null Hypothesis
H_B: Bias of the gender of users involved in fake news with respect to real	Reject Null Hypothesis
H_C: Bias of media usage in fake news with respect to real news.	Reject Null Hypothesis
H_D: Bias in the proportion of a particular gender of the user on the bias in the proportion of sentiments in fake news with respect to real news.	Reject Null Hypothesis
H_E: Bias in the proportion of a particular sentiment used in fake news between different gender of users.	Reject Null Hypothesis
H_F: Bias of inducing a particular sentiment with the usage of media in fake	Reject Null Hypothesis
H_G: Bias in the usage of media amongst different sentiments used in fake	Reject Null Hypothesis
H_H: Relationship between a particular gender and media usage in fake	Inconclusive
H_I: Bias in and usage of media in fake news between different gender of	Inconclusive
H_J: Significantly distinguishable bias of "follower" count in fake news.	Reject Null Hypothesis
H_K: Significantly distinguishable bias of "friends" count in fake news.	Reject Null Hypothesis
H_L: Significantly distinguishable bias of "status" count in fake news.	Fail to Reject Null
H_M: Significantly distinguishable bias of "retweet" count in fake news.	Reject Null Hypothesis
H_N: Significantly distinguishable bias of "favorite" count in fake news.	Fail to Reject Null

In our qualitative hypotheses H_A, it is assumed that there is a bias in the proportions of sentiment (linguistic tone) in fake news. Although, the central polarity of bias was unclear. With our study on two COVID-19 specific datasets, we found a strong bias of fake news towards neutral sentiment followed by negative sentiment with respect to real news, which is proved by the results of our first hypothesis. In the second hypothesis, H_B, we tested the bias in the proportion of gender in fake news. The results predicted that there is a strong bias of the male gender towards fake news propagation with respect to real news. Now the influence of the gender ratio of Twitter users is not taken into account as the test is performed to distinguish characteristics of real news and fake news. Any sort of this influence is assumed to affect both types of news equally and nullify its effect. In other words, the speculated gender ratio of

6.85:3.15 should be observed in any random sample collection of tweets. Hence, we directly compare the actual ratio from the dataset without considering the deviation from the speculated ratio. In our datasets, the proportion of tweets (both real and fake) with media is more minor than tweets without media. From the chi-square test results on hypothesis H_C, we find that the proportion of fake news with media is significantly less than expected and substantially more than anticipated for real news with media. Further, we explore if the bias in proportions of one category amongst sentiment, gender, and media usage, is significantly influenced by the bias in proportions of these categories. From the test for Hypothesis H_D, we find that Fake news shared by both male and female gender show bias in proportion of sentiment. The result for hypothesis H_E indicates that this bias is towards fake news being sentiment Neutral, followed by sentiment negative, with respect to real news. This supports our Hypothesis H_A. Further, from the results of testing Hypothesis H_F and H_G, H_G concludes that there is a bias of sentiment in both "with" and "without" media usage. From H_F, we conclude that this bias in fake news propagation is proportional to using positive sentiment. For the remaining combination of gender and media usage, from the results of hypotheses H_H and H_I, it cannot be concluded if there is a mutual influence of Media usage and gender of the user in the bias observed in Hypothesis H_B and H_C due to the contradictory results from the two datasets. In Hypothesis H_H, the contradictory results for "media used" and for H_I, the contradictory results for "Female" gender.

From the quantitative variables, we observe a significant distinguishable difference in the mean number of followers, friends, and retweets for fake and real news. The smaller value of mean for followers can be attributed to why most Real news sources are official media channels and celebrity users who share information on Twitter. In contrast, fake news comes mostly from regular Twitter users who do not have such a huge following. Similar reasons can be attributed to a smaller mean value for retweets of fake news. For the larger value of mean for the number of friends, we understand that the users who propagate fake news are involved in more mutual social connections. Understandably, celebrities and official media sources, when compared to active regular Twitter users, do not have many mutual connections that Twitter classifies as "friends" and, therefore, the resulting smaller value of the mean. The confidence interval for mean for each of these plots acts as a range for true mean for fake and real news and can be used to identify any sample of data by comparing its mean to the 95% CI for the mean of these plots. The non-distinguishable mean value and reverse in the plotted trend for the number of statuses posted by the users who propagated fake and real news and the

difference of range for the mean of the number of users who favorited the tweet between the two datasets make these variables unsuitable for classification of the label for the tweet.

6.6 SUMMARY

Fake information on social platforms has constantly been increasing. In the state of the COVID-19 pandemic, this problem has grown at an exponential rate globally. The pandemic is one major event generating misinformation and promoting its consumption through social networks worldwide. In the absence of a holistic fake news detection model, it is unclear what factors can be used to identify misinformation. Very few past works are dedicated to identifying such factors. In this work, we examined several factors from two Twitter datasets, MediaEval 2020 and CovidHeRA, using fourteen hypotheses HA to HN. The study uses Chi-square tests for nine qualitative theories (HA to HI), whereas for five quantitative tests (HJ to HN), we have calculated Confidence Intervals using Analysis of Means. Observations from this study unravel specific characteristics to distinguish fake news from real news. These new findings pave the way for future research and development of fake news detection algorithms. We motivate fellow researchers to design algorithms that utilize the discovered dependencies using their combined decisions. Also, we encourage to discover more identifiers that can characterize false information present online ubiquitously. This study provides a new dimension to the existing literature in the fake news domain.

CHAPTER 7

CONCLUSION

Uncontrolled and imitative information being over-loaded on the net wants applicable solutions for the complexities being generated and has become a tough nut to crack. Deep learning algorithms are proving economical and providing effective solutions with exceptional results. These solutions are to be unearthed from inconceivable horizons which too among a really precise and restricted amount because the flow of complexities has reached the verge of parallel solutions. There has been a fast increase in luring solutions for multimodal pretend news detection adopting various variant techniques. This survey permits us to conclude that deep learning architectures prove surprisingly capable of pretend news detection. they need resulted in high accuracies beneath the text-domain. repeated Neural Networks, LSTMs, GRU, Bi-directional GRU have contributed considerably to text classification. once it involves visual data, Convolutional Neural Networks kind the larger picture. Survey displays that over 40% of methodologies have incorporated CNNs and their mixtures with RNNs or different DNNs in their detection frameworks and served good results. CNNs are taking the lead in pc vision, and allied domains and became a prospective application for future FND tasks. several researchers have known fake pictures and videos and tampered regions in them, that we review as validatory tasks which will facilitate classify pretend news supported fake visuals. we tend to inspire the readers to mix such tasks with FND modules to perform multimodal FND. By fusing modules acting such tasks on totally different modalities, optimized performances are assured. There has been the inaccessibility of symbolic literature during this domain. The progress on the pathways of multimodal fake news detection has been slow. Researchers are unaware of the advancements to this point reached. Existing literature is concentrated upon fake matter news and its detection mechanisms. regarding future works, we tend to promote a multimodal framework that might with efficiency detect pretend news all told forms that revolve round the internet. we propose exploring the domain incorporating fake news within the variety of videos. we tend to additionally inspire analysisers to have interaction in building versatile multimodal datasets for future use aggregation information from websites, on-line social platforms, and therefore the likes. we tend to encourage the readers to dive deeper into machine learning and deep learning algorithms and fish out final solutions to the matter domain. Our work helps in bridging the research gaps and function potential future opportunities to figure upon. we tend to conclude this anticipating that interested researchers

can enjoy the knowledge provided and slim down their interests to the current domain to contribute to the society and analysis community.

7.1 POTENTIAL DIRECTIONS

Research done in the past years is overwhelming yet, insufficient to cope with the amount of fake news pouring in. Each new happening or event in the world serves as a topic for fake news generation and propagation. In the present scenario, while a pandemic is going on, fake news reaches out to people more swiftly than authentic news is. No data modality is left behind in the race of spreading fake news. Text is not the only type of data of which one should be aware while intaking. People need to be more careful while digesting anything available on the internet because false information could greet us in any form, be it text, image, or video. So is the need for designing efficient and robust detection mechanisms. Analyzing the limitations and research gaps, this section highlights the potential directions where research can be proceeded into.

1. **Datasets:** From the above analysis, it is obvious that there is a shortage of large-scale multimodal datasets. Machine learning and Deep learning algorithms are data-driven. With the shortage of benchmark datasets, it becomes challenging to build detection mechanisms and compare various techniques' performances. Although there are a few text datasets, those with multimodal information are limited and of poor quality. With the advantage of web-scraping mechanisms and free APIs, it has become easier to collect data. To proceed in the direction of multimodality, the collection of large-scale multimodal datasets is promoted.

2. **Real-time Detection:** With the assistance of deep learning algorithms, real-time detection models can be built to use fact-checked articles on the web for training and generate predictions for unseen data. There is a wide opportunity for the development of real-time detectors and automated fact-checkers.

3. **Early Detection:** Fake news detectors are built by feeding past data to algorithms. A baseline comparison is made on previous data. These algorithms are built when the fake information has already spread into the world and affected many. Entrapped within the fake news web, the world requires early detection of false news as and when it appears online.

Users can only be benefited from fake news detectors when they provide early detection to prevent the propagation of fake news to a large scale. Early detection would allow intervention and thus mitigation of fake news before it spreads to a larger audience.

4. **Ubiquitous Detection Model:** With many social networking platforms available, it is challenging to incorporate a fake news detection mechanism to separate platforms individually. Similar content makes rounds on multiple platforms because one user can have accounts on various networks. This creates a replica of data on different social networks. With the help of redundant data and manual annotations, classification becomes easy for deep neural networks. A cross-platform system is required that would detect fake content on multiple social platforms. Implementation of models that can train on manually annotated content on one platform and then identify fake news on other platforms is suggested.

5. **Data-oriented Detection:** As far as the previous research is considered, we have very few frameworks that provide credibility assessment to fake content types. Most techniques consider text only, while some allow visual verification. It is challenging for a single system to verify the contents of all data modalities. Such a system would be more beneficial for the general public to authenticate information.

6. **Feature-oriented Detection:** All existing approaches use a limited subset of features, either linguistics, visuals, hybrid, data-centered, sentiment scores, social context, network-based, user-based, or post-based features. These contributing factors of fake news identification could be used all together for dependable predictions.

7. **Integrity Assessment:** In multimodal approaches, existing works perform detection based on features from each type of data independently. In many fake news instances, the post contents are not semantically related. The text, image, or video for a given post could be expressing unrelated context. Few works focus on assessing the semantic integrity of the news. This helps detect false news where data modalities have not been manipulated but are unrelated to each other. Such integrity assessment tools shall help in identifying out-of-context news items.

8. **Embedded Fake News Detection:** Detection of fake embedded content has not been done yet. A large volume of fake news is spreading through such type of data. To cope up with the incoming fake news, this type of detection mechanism is required.

9. **Multilingual Detection:** Current approaches have focused on English language data in text and videos. Due to the spread of fake news through regional languages on the web, multi-lingual approaches should be considered to detect fake news from other languages in the form of text, videos, or embedded content.

10. **Data Manipulation Detection:** With the popularity of image and video forensics techniques, forgery detection in data has become easier. Various manipulation techniques like face-spoofing detection, deepfake identification, tampering detection, splicing, copy-move detection, object removal/addition detection, etc., should be considered and merged with fake news detection mechanisms. There is a need for merging the domains of fake news detection and data manipulation detection.

11. **Browser Plugin/Application Software:** The availability of fake news detector tools in the form of easy-to-use browser plugins, add-ons, software, and mobile applications will enhance their accessibility and serve detection on a user-basis.

RELATED PUBLICATIONS

[1] Raj, Chahat, and Meel, Priyanka. “ConvNet Frameworks for Multimodal Fake News Detection.” Applied Intelligence, Springer (2021), 1-8. [Published]

[2] Raj, Chahat, and Meel, Priyanka. “Microblogs Deception Detection using BERT and Multiscale CNN.” ICACECS (2021). [Accepted]

[3] Raj, Chahat, and Meel, Priyanka. “ARCNN Frameworks for Multimodal Infodemic Detection” Neural Networks (2021). [Communicated & Revised]

[4] Raj, Chahat, and Meel, Priyanka. "Fake News on Multiple Online Social Networks." ICPCCAI (2020): 1-8. [Published]

[5] Raj, Chahat, and Meel, Priyanka. “Fake News Characterization and Factor Identification: A Statistical Approach” Technology in Society (2021). [Under Review]

[6] Raj, Chahat, and Meel, Priyanka. "A Review of Web Infodemic Analysis and Detection Trends across Multimodalities using Deep Neural Networks" Neurocomputing (2021). [Under Review]

REFERENCES

[1] https://www.news18.com/news/buzz/twitter-trends-with-world-war-iii-after-killing-of-irans-top-commander-in-us-airstrike-2444295.html

[2] https://timesofindia.indiatimes.com/life-style/health-fitness/health-news/coronavirus-myth-vs-fact-whatsapp-forward-claiming-turmeric-and-black-pepper-home-remedy-to-cure-covid-19-is-fake/photostory/76995286.cms?picid=76995474

[3] https://www.pedestrian.tv/news/french-government-cocaine-coke-coronavirus-hoax/

[4] https://en.wikipedia.org/wiki/List_of_unproven_methods_against_COVID-19

[5] https://www.boomlive.in/fake-news/viral-messages-stating-who-protocols-for-lockdown-extension-are-false-7552

[6] https://www.bbc.com/news/world-52224331

[7] https://en.wikipedia.org/wiki/Pizzagate_conspiracy_theory

[8] Reis, J. C., Melo, P., Garimella, K., Almeida, J. M., Eckles, D., & Benevenuto, F. (2020, May). A Dataset of Fact-Checked Images Shared on WhatsApp During the Brazilian and Indian Elections. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 14, pp. 903-908).

[9] A. Gupta, H. Lamba, P. Kumaraguru, and A. Joshi, "Faking sandy: Characterizing and identifying fake images on twitter during hurricane sandy," *WWW 2013 Companion - Proc. 22nd Int. Conf. World Wide Web*, pp. 729–736, 2013.

[10] https://www.business-standard.com/article/current-affairs/gps-chips-and-radioactive-ink-in-new-notes-top-10-fake-news-in-2016-116122600083_1.html

[11] https://en.wikipedia.org/wiki/Fake_news_in_India#Citizenship_(Amendment)_Act_2019

[12] https://en.wikipedia.org/wiki/Fake_news_in_India#Kashmir

[13] Oshikawa, R., Qian, J., & Wang, W. Y. (2018). A survey on natural language processing for fake news detection. *arXiv preprint arXiv:1811.00770*.

[14] Ahmed, H., Traore, I., & Saad, S. (2018). Detecting opinion spams and fake news using text classification. *Security and Privacy, 1*(1), e9.

[15] Kula, S., Choraś, M., Kozik, R., Ksieniewicz, P., & Woźniak, M. (2020, June). Sentiment analysis for fake news detection by means of neural networks. In *International Conference on Computational Science* (pp. 653-666). Springer, Cham.

[16] Han, Y., Karunasekera, S., & Leckie, C. (2020). Graph Neural Networks with Continual Learning for Fake News Detection from Social Media. *arXiv preprint arXiv:2007.03316*.

[17] https://www.hindustantimes.com/more-lifestyle/covid-19-misinformation-fake-news-on-coronavirus-is-proving-to-be-contagious/story-wmKXCKjESMQIEZoyI9VJ8O.html

[18] https://www.kron4.com/news/media-reports-say-north-korean-dictator-kim-jong-un-dead/

[19] S. Sharma and D. K. Sharma, "Fake News Detection: A long way to go," *2019 4th Int. Conf. Inf. Syst. Comput. Networks, ISCON 2019*, no. May 2020, pp. 816–821, 2019.

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

[21] F. Torabi Asr and M. Taboada, "Big Data and quality data for fake news and misinformation detection," *Big Data Soc.*, vol. 6, no. 1, pp. 1–14, 2019.

[22] X. Zhang and A. A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion," *Inf. Process. Manag.*, no. February, pp. 1–26, 2019.

[23] E. C. Tandoc, Z. W. Lim, and R. Ling, “Defining ‘Fake News’: A typology of scholarly definitions,” *Digit. Journal.*, vol. 6, no. 2, pp. 137–153, 2018.
[24] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake News Detection on Social Media,” *ACM SIGKDD Explor. Newsl.*, vol. 19, no. 1, pp. 22–36, 2017.
[25] V. Mosinzova, B. Fabian, T. Ermakova, and A. Baumann, “Fake News, Conspiracies and Myth Debunking in Social Media - A Literature Survey Across Disciplines,” *SSRN Electron. J.*, no. February, 2019.
[26] V. L. Rubin, Y. Chen, and N. J. Conroy, “Deception detection for news: Three types of fakes,” *Proc. Assoc. Inf. Sci. Technol.*, vol. 52, no. 1, pp. 1–4, 2015.
[27] G. Rajendran, B. Chitturi, and P. Poornachandran, “Stance-In-Depth Deep Neural Approach to Stance Classification,” *Procedia Comput. Sci.*, vol. 132, no. Iccids, pp. 1646–1653, 2018.
[28] J. Cao, P. Qi, Q. Sheng, T. Yang, J. Guo, and J. Li, “Exploring the Role of Visual Content in Fake News Detection,” pp. 141–161, 2020.
[29] S. B. Parikh and P. K. Atrey, “Media-Rich Fake News Detection: A Survey,” *Proc. - IEEE 1st Conf. Multimed. Inf. Process. Retrieval, MIPR 2018*, no. April, pp. 436–441, 2018.
[30] Widiastuti, N. I. "Convolution Neural Network for Text Mining and Natural Language Processing." In *IOP Conference Series: Materials Science and Engineering*, vol. 662, no. 5, p. 052010. IOP Publishing, 2019.
[31] Al-Saffar, Ahmed Ali Mohammed, Hai Tao, and Mohammed Ahmed Talab. "Review of deep convolution neural network in image classification." In *2017 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET)*, pp. 26-31. IEEE, 2017.
[32] Karpathy, Andrej, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. "Large-scale video classification with convolutional neural networks." In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 1725-1732. 2014.
[33] Singh, Rahul Dev, Ajay Mittal, and Rajesh K. Bhatia. "3D convolutional neural network for object recognition: a review." *Multimedia Tools and Applications* 78, no. 12 (2019): 15951-15995.
[34] Zhao, Bendong, Huanzhang Lu, Shangfeng Chen, Junliang Liu, and Dongya Wu. "Convolutional neural networks for time series classification." *Journal of Systems Engineering and Electronics* 28, no. 1 (2017): 162-169.
[35] Kwon, Donghwoon, Kathiravan Natarajan, Sang C. Suh, Hyunjoo Kim, and Jinoh Kim. "An empirical study on network anomaly detection using convolutional neural networks." In *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, pp. 1595-1598. IEEE, 2018.
[36] Huang, Jui-Ting, Jinyu Li, and Yifan Gong. "An analysis of convolutional neural networks for speech recognition." In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4989-4993. IEEE, 2015.
[37] Ghosh, Mahmoud M. Abu, and Ashraf Y. Maghari. "A comparative study on handwriting digit recognition using neural networks." In *2017 International Conference on Promising Electronic Technologies (ICPET)*, pp. 77-81. IEEE, 2017.
[38] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, “A survey of the recent architectures of deep convolutional neural networks,” *Artif. Intell. Rev.*, pp. 1–69, 2020.
[39] G. Krishnamurthy, N. Majumder, S. Poria, and E. Cambria, “A Deep Learning Approach for Multi-modal Deception Detection,” 2018.
[40] Mou, Lichao, Pedram Ghamisi, and Xiao Xiang Zhu. "Deep recurrent neural networks for hyperspectral image classification." *IEEE Transactions on Geoscience and Remote*

Sensing 55, no. 7 (2017): 3639-3655.
[41] Yue-Hei Ng, Joe, Matthew Hausknecht, Sudheendra Vijayanarasimhan, Oriol Vinyals, Rajat Monga, and George Toderici. "Beyond short snippets: Deep networks for video classification." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4694-4702. 2015.
[42] Liang, Ming, and Xiaolin Hu. "Recurrent convolutional neural network for object recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3367-3375. 2015.
[43] Pan, Pingbo, Zhongwen Xu, Yi Yang, Fei Wu, and Yueting Zhuang. "Hierarchical recurrent neural encoder for video representation with application to captioning." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1029-1038. 2016.
[44] Madhavan, P. G. "Recurrent neural network for time series prediction." In *Proceedings of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Societ*, pp. 250-251. IEEE, 1993.
[45] Malhotra, Pankaj, Lovekesh Vig, Gautam Shroff, and Puneet Agarwal. "Long short term memory networks for anomaly detection in time series." In *Proceedings*, vol. 89, pp. 89-94. Presses universitaires de Louvain, 2015.
[46] Ruales, Joaquın. "Recurrent neural networks for sentiment analysis." *IEEE. Colombia: Colombia University* (2011).
[47] M. A. Qureshi and M. Deriche, "A bibliography of pixel-based blind image forgery detection techniques," *Signal Process. Image Commun.*, vol. 39, pp. 46–74, 2015.
[48] D. Brezeale and D. J. Cook, "Automatic video classification: A survey of the literature," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 38, no. 3, pp. 416–430, 2008.
[49] C. Boididou *et al.*, "Verifying information with multimedia content on twitter: A comparative study of automated approaches," *Multimed. Tools Appl.*, vol. 77, no. 12, pp. 15545–15571, 2018.
[50] K. Anoop, M. P. Gangan, D. P, and V. L. Lajish, *Leveraging Heterogeneous Data for Fake News Detection*. 2019.
[51] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Deepfakes and beyond: A Survey of face manipulation and fake detection," *Inf. Fusion*, vol. 64, pp. 131–148, 2020.
[52] N. Saini, M. Singhal, M. Tanwar, and P. Meel, "Multi-modal, Semi-supervised and Unsupervised web content credibility analysis Frameworks," *Proc. Int. Conf. Intell. Comput. Control Syst. ICICCS 2020*, no. Iciccs, pp. 948–955, 2020.
[53] S. Elkasrawi, A. Dengel, A. Abdelsamad, and S. S. Bukhari, "What You See is What You Get? Automatic Image Verification for Online News Content," *Proc. - 12th IAPR Int. Work. Doc. Anal. Syst. DAS 2016*, pp. 114–119, 2016.
[54] Y. Wang *et al.*, "EANN: Event adversarial neural networks for multi-modal fake news detection," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 849–857, 2018.
[55] Z. Jin, J. Cao, J. Luo, and Y. Zhang, "Image Credibility Analysis with Effective Domain Transferred Deep Networks," pp. 1–10, 2016.
[56] P. Qi, J. Cao, T. Yang, J. Guo, and J. Li, "Exploiting multi-domain visual information for fake news detection," *Proc. - IEEE Int. Conf. Data Mining, ICDM*, vol. 2019-Novem, pp. 518–527, 2019.
[57] D. K. Vishwakarma, D. Varshney, and A. Yadav, "Detection and veracity analysis of fake news via scrapping and authenticating the web search," *Cogn. Syst. Res.*, vol. 58, pp. 217–229, 2019.
[58] C. Pasquini, C. Brunetta, A. F. Vinci, V. Conotter, and G. Boato, "TOWARDS THE

VERIFICATION OF IMAGE INTEGRITY IN ONLINE NEWS Cecilia Pasquini , Carlo Brunetta , Andrea F . Vinci , Valentina Conotter , Giulia Boato," *Multimed. Expo Work. (ICMEW), 2015 IEEE Int. Conf.*, pp. 1–6, 2015.

[59] L. Self-consistency, M. Huh, A. Liu, A. Owens, A. A. Efros, and U. C. Berkeley, "Fighting Fake News : Image Splice Detection," 2018.

[60] L. Cui, S. Wang, and D. Lee, "Same: Sentiment-aware multi-modal embedding for detecting fake news," *Proc. 2019 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining, ASONAM 2019*, pp. 41–48, 2019.

[61] Z. Jin, J. Cao, Y. Zhang, and Y. Zhang, "MCG-ICT at MediaEval 2015: Verifying multimedia use with a two-level classification model," *CEUR Workshop Proc.*, vol. 1436, 2015.

[62] K. Shu, X. Zhou, S. Wang, R. Zafarani, and H. Liu, "The role of user profiles for fake news detection," *Proc. 2019 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining, ASONAM 2019*, pp. 436–439, 2019.

[63] O. Ajao, D. Bhowmik, and S. Zargari, "Fake news identification on Twitter with hybrid CNN and RNN models," *ACM Int. Conf. Proceeding Ser.*, no. July, pp. 226–230, 2018.

[64] S. Singhal, R. R. Shah, T. Chakraborty, P. Kumaraguru, and S. Satoh, "SpotFake: A multi-modal framework for fake news detection," *Proc. - 2019 IEEE 5th Int. Conf. Multimed. Big Data, BigMM 2019*, pp. 39–47, 2019.

[65] D. Saez-Trumper, "Fake tweet buster: A webtool to identify users promoting fake news ontwitter," *HT 2014 - Proc. 25th ACM Conf. Hypertext Soc. Media*, pp. 316–317, 2014.

[66] A. Gupta, H. Lamba, P. Kumaraguru, and A. Joshi, "Faking sandy: Characterizing and identifying fake images on twitter during hurricane sandy," *WWW 2013 Companion - Proc. 22nd Int. Conf. World Wide Web*, pp. 729–736, 2013.

[67] F. Lago, Q. T. Phan, G. Boato, and A. Venčkauskas, "Visual and Textual Analysis for Image Trustworthiness Assessment within Online News," *Secur. Commun. Networks*, vol. 2019, 2019.

[68] Y. Yang, L. Zheng, J. Zhang, Q. Cui, Z. Li, and P. S. Yu, "TI-CNN: Convolutional Neural Networks for Fake News Detection," 2018.

[69] S. Huckle and M. White, "Fake News: A Technological Approach to Proving the Origins of Content, Using Blockchains," *Big Data*, vol. 5, no. 4, pp. 356–371, 2017.

[70] S. Knshnan and M. Chen, "Identifying tweets with fake news," *Proc. - 2018 IEEE 19th Int. Conf. Inf. Reuse Integr. Data Sci. IRI 2018*, vol. 67, pp. 460–464, 2018.

[71] G. Armano *et al.*, "NewsVallum: Semantics-Aware Text and Image Processing for Fake News Detection system," *CEUR Workshop Proc.*, vol. 2161, 2018.

[72] X. Zhou, J. Wu, and R. Zafarani, "SAFE: Similarity-Aware Multi-modal Fake News Detection," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12085 LNAI, no. 1, pp. 354–367, 2020.

[73] Y. Chen, N. J. Conroy, and V. L. Rubin, "News in an online world: The need for an 'automatic crap detector,'" *Proc. Assoc. Inf. Sci. Technol.*, vol. 52, no. 1, pp. 1–4, 2015.

[74] E. Müller-Budack, J. Theiner, S. Diering, M. Idahl, and R. Ewerth, "Multi-modal analytics for real-world news using measures of cross-modal entity consistency," *ICMR 2020 - Proc. 2020 Int. Conf. Multimed. Retr.*, pp. 16–25, 2020.

[75] S. B. Parikh, S. R. Khedia, and P. K. Atrey, "A framework to detect fake tweet images on social media," *Proc. - 2019 IEEE 5th Int. Conf. Multimed. Big Data, BigMM 2019*, no. September, pp. 104–110, 2019.

[76] L. Nixon, E. Apostolidis, F. Markatopoulou, I. Patras, and V. Mezaris, "Multi-modal Video Annotation for Retrieval and Discovery of Newsworthy Video in a News Verification Scenario," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11295 LNCS, no. Mmm, pp. 143–155, 2019.

[77] A. Bagade *et al.*, “The Kauwa-Kaate fake news detection system: DemO,” *ACM Int. Conf. Proceeding Ser.*, pp. 302–306, 2020.
[78] C. Boididou, S. Papadopoulos, M. Zampoglou, L. Apostolidis, O. Papadopoulou, and Y. Kompatsiaris, “Detection and visualization of misleading content on Twitter,” *Int. J. Multimed. Inf. Retr.*, vol. 7, no. 1, pp. 71–86, 2018.
[79] S. Sun, H. Liu, J. He, and X. Du, “Detecting Event Rumors on Sina Weibo Automatically,” pp. 120–131, 2013.
[80] B. Bayar and M. C. Stamm, “A deep learning approach to universal image manipulation detection using a new convolutional layer,” *IH MMSec 2016 - Proc. 2016 ACM Inf. Hiding Multimed. Secur. Work.*, pp. 5–10, 2016.
[81] E. Sabir, Y. Wu, W. A. Almageed, and P. Natarajan, “Deep multi-modal image-repurposing detection,” *MM 2018 - Proc. 2018 ACM Multimed. Conf.*, vol. 2, pp. 1337–1345, 2018.
[82] A. Jaiswal, Y. Wu, W. Abdalmageed, I. Masi, and P. Natarajan, “Aird: Adversarial learning framework for image repurposing detection,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 11322–11331, 2019.
[83] T. Pomari, G. Ruppert, E. Rezende, A. Rocha, and T. Carvalho, “Image Splicing Detection Through Illumination Inconsistencies and Deep Learning,” *Proc. - Int. Conf. Image Process. ICIP*, no. September, pp. 3788–3792, 2018.
[84] M. Zampoglou, S. Papadopoulos, and Y. Kompatsiaris, “DETECTING IMAGE SPLICING IN THE WILD (WEB) Markos Zampoglou , Symeon Papadopoulos , Yiannis Kompatsiaris,” *2015 IEEE Int. Conf. Multimed. Expo Work.*, pp. 1–6.
[85] Y. Wu, W. Abd-Almageed, and P. Natarajan, “Deep matching and validation network: An end-to-end solution to constrained image splicing localization and detection,” *MM 2017 - Proc. 2017 ACM Multimed. Conf.*, pp. 1480–1502, 2017.
[86] S. Tariq, S. Lee, H. Kim, Y. Shin, and S. S. Woo, “Detecting both machine and human created fake face images in the wild,” *Proc. ACM Conf. Comput. Commun. Secur.*, pp. 81–87, 2018.
[87] F. Marra, D. Gragnaniello, D. Cozzolino, and L. Verdoliva, “Detection of GAN-Generated Fake Images over Social Networks,” *Proc. - IEEE 1st Conf. Multimed. Inf. Process. Retrieval, MIPR 2018*, pp. 384–389, 2018.
[88] H. H. Nguyen, N. D. T. Tieu, H. Q. Nguyen-Son, V. Nozick, J. Yamagishi, and I. Echizen, “Modular convolutional neural network for discriminating between computer-generated images and photographic images,” *ACM Int. Conf. Proceeding Ser.*, 2018.
[89] N. Rahmouni *et al.*, “Using Convolution Neural Networks To cite this version : HAL Id : hal-01664590 Distinguishing Computer Graphics from Natural Images Using Convolution Neural Networks,” 2017.
[90] E. R. S. de Rezende, G. C. S. Ruppert, A. Theóphilo, E. K. Tokuda, and T. Carvalho, “Exposing computer generated images by using deep convolutional neural networks,” *Signal Process. Image Commun.*, vol. 66, pp. 113–126, 2018.
[91] E. Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, I. Masi, and P. Natarajan, “Recurrent Convolutional Strategies for Face Manipulation Detection in Videos,” pp. 80–87, 2019.
[92] Y. Zhang, L. Zheng, and V. L. L. Thing, “Automated face swapping and its detection,” *2017 IEEE 2nd Int. Conf. Signal Image Process. ICSIP 2017*, vol. 2017-Janua, no. August, pp. 15–19, 2017.
[93] P. Zhou, X. Han, V. I. Morariu, and L. S. Davis, “Two-Stream Neural Networks for Tampered Face Detection,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2017-July, pp. 1831–1839, 2017.
[94] L. M. Dang, S. I. Hassan, S. Im, and H. Moon, “Face image manipulation detection based on a convolutional neural network,” *Expert Syst. Appl.*, vol. 129, pp. 156–168,

2019.
[95] Y. Wu, W. Abdalmageed, and P. Natarajan, “Mantra-net: Manipulation tracing network for detection and localization of image forgeries with anomalous features,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 9535–9544, 2019.
[96] S. Y. Wang, O. Wang, R. Zhang, A. Owens, and A. Efros, “Detecting photoshopped faces by scripting photoshop,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2019-Octob, pp. 10071–10080, 2019.
[97] Y. Wu, W. Abd-Almageed, and P. Natarajan, “BusterNet: Detecting copy-move image forgery with source/target localization,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11210 LNCS, pp. 170–186, 2018.
[98] H. Li, B. Li, S. Tan, and J. Huang, “Identification of deep network generated images using disparities in color components,” *Signal Processing*, vol. 174, pp. 1–13, 2020.
[99] L. Nataraj *et al.*, “Detecting GAN generated Fake Images using Co-occurrence Matrices,” *IS T Int. Symp. Electron. Imaging Sci. Technol.*, vol. 2019, no. 5, pp. 1–7, 2019.
[100] M. Steinebach, K. Gotkowski, and H. Liu, “Fake news detection by image montage recognition,” *J. Cyber Secur. Mobil.*, vol. 9, no. 2, pp. 175–202, 2020.
[101] P. Korshunov and S. Marcel, “DeepFakes: a New Threat to Face Recognition? Assessment and Detection,” pp. 1–5, 2018.
[102] D. Guera and E. J. Delp, “Deepfake Video Detection Using Recurrent Neural Networks,” *Proc. AVSS 2018 - 2018 15th IEEE Int. Conf. Adv. Video Signal-Based Surveill.*, 2019.
[103] H. H. Nguyen, J. Yamagishi, and I. Echizen, “Capsule-forensics: Using Capsule Networks to Detect Forged Images and Videos,” *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2019-May, pp. 2307–2311, 2019.
[104] J. Yang, Z. Lei, and S. Z. Li, “Learn Convolutional Neural Network for Face Anti-Spoofing,” 2014.
[105] P. Korshunov and S. Marcel, “Speaker inconsistency detection in tampered video,” *Eur. Signal Process. Conf.*, vol. 2018-Septe, no. September, pp. 2375–2379, 2018.
[106] Z. Wu, B. Singh, L. S. Davis, and V. S. Subrahmanian, “Deception detection in videos,” *32nd AAAI Conf. Artif. Intell. AAAI 2018*, pp. 1695–1702, 2018.
[107] V. Pérez-Rosas, M. Abouelenien, R. Mihalcea, and M. Burzo, “Deception detection using real-life trial data,” *ICMI 2015 - Proc. 2015 ACM Int. Conf. Multi-modal Interact.*, pp. 59–66, 2015.
[108] Y. Li, M. C. Chang, and S. Lyu, “In Ictu Oculi: Exposing AI created fake videos by detecting eye blinking,” *10th IEEE Int. Work. Inf. Forensics Secur. WIFS 2018*, 2019.
[109] P. Bestagini, S. Milani, M. Tagliasacchi, and S. Tubaro, “Local tampering detection in video sequences,” *2013 IEEE Int. Work. Multimed. Signal Process. MMSP 2013*, pp. 488–493, 2013.
[110] Z. Jin, J. Cao, Y. Zhang, J. Zhou, and Q. Tian, “Novel Visual and Statistical Image Features for Microblogs News Verification,” *IEEE Trans. Multimed.*, vol. 19, no. 3, pp. 598–608, 2017.
[111] D. Khattar, M. Gupta, J. S. Goud, and V. Varma, “MvaE: Multi-modal variational autoencoder for fake news detection,” *Web Conf. 2019 - Proc. World Wide Web Conf. WWW 2019*, pp. 2915–2921, 2019.
[112] Papadopoulou, O., Zampoglou, M., Papadopoulos, S., & Kompatsiaris, Y. (2017, June). Web video verification using contextual cues. In *Proceedings of the 2nd International Workshop on Multimedia Forensics and Security* (pp. 6-10).

[113] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, “Multi-modal fusion with recurrent neural networks for rumor detection on microblogs,” *MM 2017 - Proc. 2017 ACM Multimed. Conf.*, pp. 795–816, 2017.

[114] Cédric Maigrot, Vincent Claveau, Ewa Kijak, Ronan Sicre. MediaEval 2016: A multi-modal system
for the Verifying Multimedia Use task. MediaEval 2016: ”Verfiying Multimedia Use” task, Oct 2016,
Hilversum, Netherlands.

[115] A. Jaiswal, E. Sabir, W. Abd Almageed, and P. Natarajan, “Multimedia semantic integrity assessment using joint embedding of images and text,” *MM 2017 - Proc. 2017 ACM Multimed. Conf.*, pp. 1465–1471, 2017.

[116] D. Zlatkova, P. Nakov, and I. Koychev, “Fact-checking meets fauxtography: Verifying claims about images,” *EMNLP-IJCNLP 2019 - 2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf.*, pp. 2099–2108, 2020.

[117] V. K. Singh, I. Ghosh, and D. Sonagara, “Detecting fake news stories via multi-modal analysis,” *J. Assoc. Inf. Sci. Technol.*, no. February, pp. 1–15, 2020.

[118] V. V Kniaz, V. Knyaz, and F. Remondino, “The Point Where Reality Meets Fantasy: Mixed Adversarial Generators for Image Splice Detection,” *Adv. Neural Inf. Process. Syst. 32*, no. NeurIPS, pp. 215–226, 2019.

[119] Nakamura, K., Levy, S., & Wang, W. Y. (2019). r/fakeddit: A new multi-modal benchmark dataset for fine-grained fake news detection. *arXiv preprint arXiv:1911.03854*.

[120] Jindal, S., Sood, R., Singh, R., Vatsa, M., & Chakraborty, T. NewsBag: A Multi-modal Benchmark Dataset for Fake News Detection.

[121] Zhou, X., Mulay, A., Ferrara, E., & Zafarani, R. (2020). ReCOVery: A Multi-modal Repository for COVID-19 News Credibility Research. *arXiv preprint arXiv:2006.05557*.

[122] Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., ... & Gao, J. (2018, July). Eann: Event adversarial neural networks for multi-modal fake news detection. *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining* (pp. 849-857).

[123] Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, *31*(2), 211-36.

[124] Burkhardt, J. M. (2017). How fake news spreads. *Library technology reports*, *53*(8), 10-13.

[125] Adiba, F. I., Islam, T., Kaiser, M. S., Mahmud, M., & Rahman, M. A. (2020). Effect of Corpora on Classification of Fake News using Naive Bayes Classifier. *International Journal of Automation, Artificial Intelligence and Machine Learning*, *1*(1), 80-92.

[126] Meel, P., & Vishwakarma, D. K. (2021). HAN, image captioning, and forensics ensemble multimodal fake news detection. *Information Sciences*, *567*, 23-41.

[127] Singh, M., Kaur, R., & Iyengar, S. R. S. (2020, December). Multidimensional Analysis of Fake News Spreaders on Twitter. In *International Conference on Computational Data and Social Networks* (pp. 354-365). Springer, Cham.

[128] Ferrara, E., Cresci, S., & Luceri, L. (2020). Misinformation, manipulation, and abuse on social media in the era of COVID-19. *Journal of Computational Social Science*, *3*(2), 271-277.

[129] Figueira, Á., & Oliveira, L. (2017). The current state of fake news: challenges and opportunities. *Procedia Computer Science*, *121*, 817-825.

[130] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, *19*(1), 22-36.
[131] Meel, P., & Vishwakarma, D. K. (2020). Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. *Expert Systems with Applications*, *153*, 112986.
[133] Narwal, B. (2018, October). Fake news in digital media. In *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)* (pp. 977-981). IEEE.
[134] Orso, D., Federici, N., Copetti, R., Vetrugno, L., & Bove, T. (2020). Infodemic and the spread of fake news in the COVID-19-era. *European Journal of Emergency Medicine*.
[135] Allahverdipour, H. (2020). Global challenge of health communication: infodemia in the coronavirus disease (COVID-19) pandemic. *J Educ Community Health*, *7*(2), 65-67.
[136] Naeem, S. B., Bhatti, R., & Khan, A. (2020). An exploration of how fake news is taking over social media and putting public health at risk. *Health Information & Libraries Journal*.
[137] Khattar, D., Goud, J. S., Gupta, M., & Varma, V. (2019, May). Mvae: Multi-modal variational autoencoder for fake news detection. In *The World Wide Web Conference* (pp. 2915-2921).
[138] Ajao, O., Bhowmik, D., & Zargari, S. (2018, July). Fake news identification on twitter with hybrid cnn and rnn models. In *Proceedings of the 9th international conference on social media and society* (pp. 226-230).
[139] Singhal, S., Shah, R. R., Chakraborty, T., Kumaraguru, P., & Satoh, S. I. (2019, September). SpotFake: A Multi-modal Framework for Fake News Detection. In *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)* (pp. 39-47). IEEE.
[140] Atrey, P. K., Hossain, M. A., El Saddik, A., & Kankanhalli, M. S. (2010). Multimodal fusion for multimedia analysis: a survey. *Multimedia systems*, *16*(6), 345-379.
[141] Majumder, S. B., & Das, D. (2020). Detecting Fake News Spreaders on Twitter Using Universal Sentence Encoder. In *CLEF*.
[142] Mookdarsanit, P., & Mookdarsanit, L. (2021). The COVID-19 fake news detection in Thai social texts. *Bulletin of Electrical Engineering and Informatics*, *10*(2), 988-998.
[143] Elhadad, M. K., Li, K. F., & Gebali, F. (2020, August). COVID-19-FAKES: A twitter (Arabic/English) dataset for detecting misleading information on COVID-19. In *International Conference on Intelligent Networking and Collaborative Systems* (pp. 256-268). Springer, Cham.
[144] Elhadad, M. K., Li, K. F., & Gebali, F. (2020, August). An Ensemble Deep Learning Technique to Detect COVID-19 Misleading Information. In *International Conference on Network-Based Information Systems* (pp. 163-175). Springer, Cham.
[145] Al-Rakhami, M. S., & Al-Amri, A. M. (2020). Lies Kill, Facts Save: Detecting COVID-19 Misinformation in Twitter. *IEEE Access*, *8*, 155961-155970.
[146] Al-Ahmad, B., Al-Zoubi, A. M., Abu Khurma, R., & Aljarah, I. (2021). An Evolutionary Fake News Detection Method for COVID-19 Pandemic Information. *Symmetry*, *13*(6), 1091.
[147] Shim, J. S., Lee, Y., & Ahn, H. (2021). A Link2vec-based Fake News Detection Model using Web Search Results. *Expert Systems with Applications*, 115491.
[148] Kaliyar, R. K., Goswami, A., & Narang, P. (2021). A Hybrid Model for Effective Fake News Detection with a Novel COVID-19 Dataset. In *ICAART (2)* (pp. 1066-1072).

[149] Vishwakarma, D. K., & Jain, C. (2020, June). Recent State-of-the-art of Fake News Detection: A Review. In *2020 International Conference for Emerging Technology (INCET)* (pp. 1-6). IEEE.
[150] Boididou, C., Middleton, S. E., Jin, Z., Papadopoulos, S., Dang-Nguyen, D. T., Boato, G., & Kompatsiaris, Y. (2018). Verifying information with multimedia content on twitter. *Multimedia Tools and Applications*, *77*(12), 15545-15571.
[151] Anoop, K., Gangan, M. P., Deepak, P., & Lajish, V. L. (2019). Leveraging heterogeneous data for fake news detection. In *Linking and Mining Heterogeneous and Multi-view Data* (pp. 229-264). Springer, Cham.
[152] Saini, N., Singhal, M., Tanwar, M., & Meel, P. (2020, May). Multi-modal, Semi-supervised and Unsupervised web content credibility analysis Frameworks. In *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 948-955). IEEE.
[153] Singh, V. K., Ghosh, I., & Sonagara, D. (2020). Detecting fake news stories via multi-modal analysis. *Journal of the Association for Information Science and Technology*.
[154] Yang, Y., Zheng, L., Zhang, J., Cui, Q., Li, Z., & Yu, P. S. (2018). TI-CNN: Convolutional neural networks for fake news detection. *arXiv preprint arXiv:1806.00749*.
[155] Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., ... & Gao, J. (2018, July). Eann: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining* (pp. 849-857).
[156] Cui, L., Wang, S., & Lee, D. (2019, August). SAME: sentiment-aware multi-modal embedding for detecting fake news. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 41-48).
[157] Jin, Z., Cao, J., Zhang, Y., & Zhang, Y. (2015, September). MCG-ICT at MediaEval 2015: Verifying Multimedia Use with a Two-Level Classification Model. In *MediaEval*.
[158] Shu, K., Zhou, X., Wang, S., Zafarani, R., & Liu, H. (2019, August). The role of user profiles for fake news detection. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 436-439).
[159] Krishnamurthy, G., Majumder, N., Poria, S., & Cambria, E. (2018). A deep learning approach for multi-modal deception detection. *arXiv preprint arXiv:1803.00344*.
[160] Lago, F., Phan, Q. T., & Boato, G. (2019). Visual and Textual Analysis for Image Trustworthiness Assessment within Online News. *Security and Communication Networks*, *2019*.
[161] Maigrot, C., Claveau, V., Kijak, E., & Sicre, R. (2016, October). Mediaeval 2016: A multi-modal system for the verifying multimedia use task.
[162] Shahi, G. K., & Nandini, D. (2020). FakeCovid--A Multilingual Cross-domain Fact Check News Dataset for COVID-19. *arXiv preprint arXiv:2006.11343*.
[163] Xiao, L., Wang, G., & Zuo, Y. (2018, December). Research on patent text classification based on word2vec and LSTM. In *2018 11th International Symposium on Computational Intelligence and Design (ISCID)* (Vol. 1, pp. 71-74). IEEE.
[164] Zhou, X., Mulay, A., Ferrara, E., & Zafarani, R. (2020). ReCOVery: A Multi-modal Repository for COVID-19 News Credibility Research. *arXiv preprint arXiv:2006.05557*.
[165] Cui, L., & Lee, D. (2020). CoAID: COVID-19 Healthcare Misinformation Dataset. *arXiv preprint arXiv:2006.00885*.

[166] Pogorelov, K., Schroeder, D. T., Burchard, L., Moe, J., Brenner, S., Filkukova, P., & Langguth, J. (2020, December). Fakenews: Corona virus and 5g conspiracy task at mediaeval 2020. In *MediaEval 2020 Workshop*.
[167] Jin, Z., Cao, J., Guo, H., Zhang, Y., & Luo, J. (2017, October). Multi-modal fusion with recurrent neural networks for rumor detection on microblogs. In *Proceedings of the 25th ACM international conference on Multimedia* (pp. 795-816).
[168] Soltaninejad, K. (2020). Methanol mass poisoning outbreak, a consequence of COVID-19 pandemic and misleading messages on social media. The international journal of occupational and environmental medicine, 11(3), 148.
[169] Tagliabue, F., Galassi, L., & Mariani, P. (2020). The "pandemic" of disinformation in COVID-19. SN comprehensive clinical medicine, 2(9), 1287-1289.
[170] Zannettou, S., Bradlyn, B., De Cristofaro, E., Kwak, H., Sirivianos, M., Stringini, G., & Blackburn, J. (2018, April). What is gab: A bastion of free speech or an alt-right echo chamber. In Companion Proceedings of the The Web Conference 2018 (pp. 1007-1014).
[171] Altay, S., Hacquin, A. S., & Mercier, H. (2019). Why do so few people share fake news? It hurts their reputation. new media & society, 1461444820969893.
[172] Ardèvol-Abreu, A., Delponti, P., & Rodríguez-Wangüemert, C. (2020). Intentional or inadvertent fake news sharing? Fact-checking warnings and users' interaction with social media content. Profesional de la Información, 29(5).
[173] Talwar, S., Dhir, A., Kaur, P., Zafar, N., & Alrasheedy, M. (2019). Why do people share fake news? Associations between the dark side of social media use and fake news sharing behavior. Journal of Retailing and Consumer Services, 51, 72-82.
[174] Rampersad, G., & Althiyabi, T. (2020). Fake news: Acceptance by demographics and culture on social media. Journal of Information Technology & Politics, 17(1), 1-11.
[175] Meel, P., & Vishwakarma, D. K. (2020). Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. Expert Systems with Applications, 153, 112986.
[176] Vishwakarma, D. K., & Jain, C. (2020, June). Recent State-of-the-art of Fake News Detection: A Review. In 2020 International Conference for Emerging Technology (INCET) (pp. 1-6). IEEE.
[177] Pogorelov, K., Schroeder, D. T., Burchard, L., Moe, J., Brenner, S., Filkukova, P., & Langguth, J. (2020). Fakenews: Corona virus and 5g conspiracy task at mediaeval 2020. In MediaEval 2020 Workshop.
[178] Dharawat, A., Lourentzou, I., Morales, A., & Zhai, C. (2020). Drink bleach or do what now? Covid-HeRA: A dataset for risk-informed health decision making in the presence of COVID19 misinformation. arXiv preprint arXiv:2010.08743.
[179] Parikh, S. B., Patil, V., & Atrey, P. K. (2019, March). On the origin, proliferation and tone of fake news. In 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) (pp. 135-140). IEEE.
[180] Lozano, M. G., Brynielsson, J., Franke, U., Rosell, M., Tjörnhammar, E., Varga, S., & Vlassov, V. (2020). Veracity assessment of online data. Decision Support Systems, 129, 113132.
[181] Narwal, B. (2018, October). Fake news in digital media. In 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN) (pp. 977-981). IEEE.
[182] Oshikawa, R., Qian, J., & Wang, W. Y. (2018). A survey on natural language processing for fake news detection. arXiv preprint arXiv:1811.00770.

[183] Parikh, S. B., & Atrey, P. K. (2018, April). Media-rich fake news detection: A survey. In 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) (pp. 436-441). IEEE.
[184] Zhou, X., & Zafarani, R. (2019). Network-based fake news detection: A pattern-driven approach. ACM SIGKDD Explorations Newsletter, 21(2), 48-60.
[185] Jang, S. M., & Kim, J. K. (2018). Third person effects of fake news: Fake news regulation and media literacy interventions. Computers in human behavior, 80, 295-302.
[186] Talwar, S., Dhir, A., Singh, D., Virk, G. S., & Salo, J. (2020). Sharing of fake news on social media: Application of the honeycomb framework and the third-person effect hypothesis. Journal of Retailing and Consumer Services, 57, 102197.
[187] Brewer, P. R., Young, D. G., & Morreale, M. (2013). The impact of real news about "fake news": Intertextual processes and political satire. International Journal of Public Opinion Research, 25(3), 323-343.
[188] Horne, B., & Adali, S. (2017, May). This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 11, No. 1).
[189] C. Silverman and J. Singer-Vine, "Most americans who see fake news believe it, new survey says," BuzzFeed News, 2016.

Appendix

Appendix A1: Results obtained on D1 (CovID I)

Fusion	Model	F1	Accuracy	Precision	Recall	FPR	ROC	Spec.	MCC
Early Fusion	M1	0.8145	0.8343	0.8604	0.7733	0.1115	0.9119	0.8885	0.6686
	M2	0.8477	0.8943	0.7803	0.9279	0.1213	0.9581	0.8787	0.7745
	M3	0.8049	0.8629	0.75	0.8684	0.1398	0.9249	0.8602	0.7045
	M4	0.8661	0.9029	0.8333	0.9016	0.0965	0.9588	0.9035	0.7916
	M5	0.8315	0.8657	0.8788	0.7891	0.0788	0.9427	0.9212	0.7233
	M6	0.7739	0.8286	0.6937	0.875	0.1948	0.8286	0.8052	0.65
	M7	0.7734	0.8343	0.75	0.7984	0.146	0.9176	0.854	0.6438
	M8	0.7612	0.8029	0.8333	0.7006	0.114	0.9034	0.886	0.602
	M9	0.8988	**0.9286**	0.8409	0.9652	0.0894	**0.9625**	0.9106	0.8488
	M10	0.8327	0.8657	0.8864	0.7852	0.0746	0.9388	0.9254	0.7249
Average Fusion	M1	0.8008	0.8296	0.6849	0.9638	0.2444	0.8296	0.7556	0.6887
	M2	0.8632	0.8886	0.8662	0.8601	0.0918	0.885	0.9082	0.7692
	M3	0.7489	0.8314	0.6197	0.9462	0.2101	0.7978	0.7899	0.6622
	M4	0.8955	0.92	0.8451	0.9524	0.0982	0.9081	0.9018	0.835
	M5	0.8529	0.8857	0.8169	0.8923	0.1182	0.8748	0.8818	0.7618
	M6	0.8595	0.8748	0.7658	0.9793	0.1923	0.8748	0.8077	0.7681
	M7	0.8429	0.8829	0.8271	0.8594	0.1036	0.8721	0.8964	0.7499
	M8	0.7459	0.8229	0.6842	0.8198	0.1757	0.796	0.8243	0.6175
	M9	0.8178	0.8714	0.7594	0.886	0.1356	0.8497	0.8644	0.7245
	M10	0.8095	0.8629	0.7669	0.8571	0.1342	0.8443	0.8658	0.7056
Max Fusion	M1	0.6765	0.7532	0.516	0.9817	0.3283	0.7532	0.6717	0.5753
	M2	0.7232	0.8229	0.5704	**0.9878**	0.2276	0.7828	0.7724	0.6557
	M3	0.6087	0.7686	0.4437	0.9692	0.2772	0.717	0.7228	0.5481
	M4	0.7032	0.8143	0.5423	1	0.2381	0.7711	0.7619	0.6428
	M5	0.682	0.8029	0.5211	0.9867	0.2473	0.7582	0.7527	0.6179
	M6	0.7255	0.7818	0.5766	0.978	0.3002	0.7818	0.6998	0.618
	M7	0.7611	0.8457	0.6466	0.9247	0.1829	0.8072	0.8171	0.6751
	M8	0.6763	0.8086	0.5263	0.9459	0.2283	0.7539	0.7717	0.6037
	M9	0.729	0.8343	0.5865	0.963	0.2045	0.7863	0.7955	0.6591
	M10	0.7315	0.8343	0.594	0.9518	0.2022	0.7878	0.7978	0.6568
Sum Fusion	M1	0.6725	0.7509	0.5114	0.9816	0.3304	0.7509	0.6696	0.5716
	M2	0.7232	0.8229	0.5704	0.9878	0.2276	0.7828	0.7724	0.6557
	M3	0.6087	0.7686	0.4437	0.9692	0.2772	0.717	0.7228	0.5481
	M4	0.7032	0.8143	0.5423	1	0.2381	0.7711	0.7619	0.6428
	M5	0.682	0.8029	0.5211	0.9867	0.2473	0.7582	0.7527	0.6179
	M6	0.7087	0.7722	0.5541	0.9828	0.3105	0.7722	0.6895	0.605
	M7	0.7611	0.8457	0.6466	0.9247	0.1829	0.8072	0.8171	0.6751
	M8	0.6763	0.8086	0.5263	0.9459	0.2283	0.7539	0.7717	0.6037
	M9	0.729	0.8343	0.5865	0.963	0.2045	0.7863	0.7955	0.6591
	M10	0.7315	0.8343	0.594	0.9518	0.2022	0.7878	0.7978	0.6568
Weighted Average	M1	0.8871	0.8885	0.8767	0.8978	0.1205	0.8885	0.8795	0.7772
	M2	0.8614	0.8943	0.8099	0.92	0.12	0.8809	0.88	0.7807
	M3	0.7782	0.8486	0.6549	0.9588	0.1937	0.8178	0.8063	0.6974
	M4	0.8955	0.92	0.8451	0.9524	0.0982	0.9081	0.9018	0.835
	M5	0.8571	0.8886	0.8239	0.8931	0.1142	0.8783	0.8858	0.7677
	M6	**0.8996**	0.9005	0.8919	0.9075	0.1063	0.9005	0.8937	0.8011
	M7	0.8664	0.8943	**0.9023**	0.8333	**0.0631**	0.8958	0.9369	0.7809
	M8	0.8401	0.8771	0.8496	0.8309	0.0935	0.8702	0.9065	0.7405
	M9	0.8593	0.8943	0.8496	0.8692	0.0909	0.8856	0.9091	0.7748
	M10	0.8476	0.8829	0.8571	0.8382	0.0888	0.8779	0.9112	0.7526

Appendix A2: Results obtained on D2 (CovID II)

Fusion	Model	F1	Accuracy	Precision	Recall	FPR	ROC	Spec	MCC
Early Fusion	M1	1	1	1	1	0	1	1	1
	M2	0.9124	0.9355	0.8389	1	0.0972	0.9991	0.9028	0.8703
	M3	0.9932	0.9946	0.9866	1	0.0089	1	0.9911	0.9888
	M4	0.9544	0.9651	0.9128	1	0.0551	0.9999	0.9449	0.9287
	M5	0.9829	0.9866	0.9664	1	0.0219	0.9999	0.9781	0.9722
	M6	0	0.6254	0	0	0.3746	1	0.6254	0
	M7	0.9544	0.9651	0.9128	1	0.0551	0.9945	0.9449	0.9287
	M8	0.614	0.7769	0.443	1	0.2712	0.9999	0.7288	0.5682
	M9	0.8405	0.8898	0.7248	1	0.1553	0.9999	0.8447	0.7825
	M10	0.9864	0.9892	0.9732	1	0.0176	1	0.9824	0.9778
Average Fusion	M1	0.9951	0.9951	0.9902	1	0.0097	0.995	0.9903	0.9902
	M2	0.7536	0.8629	0.6047	1	0.1735	0.8023	0.8265	0.7069
	M3	0.783	0.8763	0.6434	1	0.1592	0.8217	0.8408	0.7355
	M4	0.843	0.9059	0.7287	1	0.1259	0.8643	0.8741	0.7981
	M5	0.8684	0.9194	0.7674	1	0.1099	0.8837	0.8901	0.8265
	M6	0.9951	0.9951	0.9902	1	0.0097	0.9951	0.9903	0.9902
	M7	0.7826	0.8522	0.6644	0.9519	0.1866	0.821	0.8134	0.7019
	M8	0.7623	0.8306	0.6779	0.8707	0.1875	0.8053	0.8125	0.6459
	M9	0.8487	0.8898	0.7718	0.9426	0.136	0.8702	0.864	0.7728
	M10	0.8284	0.8763	0.745	0.9328	0.1502	0.8545	0.8498	0.7449
Max Fusion	M1	0.807	0.8383	0.6765	1	0.2444	0.8382	0.7556	0.7149
	M2	0.7228	0.8495	0.5659	1	0.1873	0.7829	0.8127	0.6782
	M3	0.7656	0.8683	0.6202	1	0.1678	0.8101	0.8322	0.7184
	M4	0.8219	0.8952	0.6977	1	0.1383	0.8488	0.8617	0.7754
	M5	0.8482	0.9086	0.7364	1	0.1227	0.8682	0.8773	0.8038
	M6	0.7965	0.8309	0.6618	1	0.2527	0.8309	0.7473	0.7032
	M7	0.7984	0.8656	0.6644	1	0.1832	0.8322	0.8168	0.7367
	M8	0.808	0.871	0.6779	1	0.1771	0.8389	0.8229	0.7469
	M9	0.8712	0.9086	0.7718	1	0.1323	0.8859	0.8677	0.8184
	M10	0.8538	0.8978	0.745	1	0.1456	0.8725	0.8544	0.7978
Sum Fusion	M1	0.807	0.8383	0.6765	1	0.2444	0.8382	0.7556	0.7149
	M2	0.7228	0.8495	0.5659	1	0.1873	0.7829	0.8127	0.6782
	M3	0.7656	0.8683	0.6202	1	0.1678	0.8101	0.8322	0.7184
	M4	0.8219	0.8952	0.6977	1	0.1383	0.8488	0.8617	0.7754
	M5	0.8482	0.9086	0.7364	1	0.1227	0.8682	0.8773	0.8038
	M6	0.7965	0.8309	0.6618	1	0.2527	0.8309	0.7473	0.7032
	M7	0.7984	0.8656	0.6644	1	0.1832	0.8322	0.8168	0.7367
	M8	0.808	0.871	0.6779	1	0.1771	0.8389	0.8229	0.7469
	M9	0.8712	0.9086	0.7718	1	0.1323	0.8859	0.8677	0.8184
	M10	0.8538	0.8978	0.745	1	0.1456	0.8725	0.8544	0.7978
Weighted Average	M1	0.9975	0.9976	0.9951	1	0.0049	0.9975	0.9951	0.9951
	M2	0.9762	0.9839	0.9535	1	0.0241	0.9767	0.9759	0.9646
	M3	0.9721	0.9812	0.9457	1	0.028	0.9729	0.972	0.9588
	M4	0.9762	0.9839	0.9535	1	0.0241	0.9767	0.9759	0.9646
	M5	0.9762	0.9839	0.9535	1	0.0241	0.9767	0.9759	0.9646
	M6	1	1	1	1	0	1	1	1
	M7	1	1	1	1	0	1	1	1
	M8	1	1	1	1	0	1	1	1
	M9	1	1	1	1	0	1	1	1
	M10	1	1	1	1	0	1	1	1

Appendix A3: Results obtained on D3 (ReCOVery news articles)

Fusion	Model	F1	Accuracy	Precision	Recall	FPR	ROC	Spec	MCC
Early Fusion	M1	0.5164	0.8012	0.3901	0.7639	0.1928	0.6725	0.8072	0.4439
	M2	0.4774	0.7666	0.4022	0.5873	0.1937	0.721	0.8063	0.3438
	M3	0.4098	0.7925	0.2717	0.8333	0.2114	0.7588	0.7886	0.396
	M4	0.4615	0.7781	0.3587	0.6471	0.1993	0.7455	0.8007	0.3592
	M5	0.4394	0.7867	0.3152	0.725	0.2052	0.7867	0.7948	0.3761
	M6	0.5887	0.8166	0.4823	0.7556	0.1706	0.712	0.8294	0.498
	M7	0.3972	0.755	0.3043	0.5714	0.2148	0.7179	0.7852	0.2814
	M8	0.56	0.8098	0.4565	0.7241	0.173	0.789	0.827	0.4659
	M9	0.4397	0.771	0.337	0.6327	0.2061	0.727	0.7939	0.3367
	M10	0.3276	0.7752	0.2065	0.7917	0.226	0.7569	0.774	0.3252
Average Fusion	M1	0.5164	0.8012	0.3901	0.7639	0.1928	0.6725	0.8072	0.4439
	M2	0.4409	0.7948	0.3733	0.5385	0.1599	0.6424	0.8401	0.3284
	M3	0.0632	0.7428	0.0333	0.6	0.2551	0.5128	0.7449	0.0938
	M4	0	0.737	0	0	0.2609	0.498	0.7391	-0.0319
	M5	0.3932	0.7948	0.2556	0.8519	0.21	0.62	0.79	0.3924
	M6	0.5887	0.8166	0.4823	0.7556	0.1706	0.712	0.8294	0.498
	M7	0.3731	0.7579	0.3205	0.4464	0.1821	0.6026	0.8179	0.2329
	M8	0.3714	0.7464	0.3333	0.4194	0.1825	0.5998	0.8175	0.2174
	M9	0.3967	0.7896	0.3077	0.5581	0.1776	0.6185	0.8224	0.3003
	M10	0.3529	0.7781	0.2692	0.5122	0.1863	0.5974	0.8137	0.252
Max Fusion	M1	0.25	0.7568	0.1489	0.7778	0.2444	0.5665	0.7556	0.2664
	M2	0.0519	0.789	0.0267	1	0.2122	0.5133	0.7878	0.1449
	M3	0	0.7399	0	0	0.2601	0.5	0.7399	0
	M4	0	0.7399	0	0	0.2601	0.5	0.7399	0
	M5	0	0.7399	0	0	0.2601	0.5	0.7399	0
	M6	0.32	0.7703	0.1986	0.8235	0.2335	0.5913	0.7665	0.3283
	M7	0.3478	0.7839	0.2564	0.5405	0.1871	0.5966	0.8129	0.2613
	M8	0.3768	0.7522	0.2559	0.4333	0.1812	0.6035	0.8188	0.2284
	M9	0.3238	0.7954	0.2179	0.6296	0.1906	0.5904	0.8094	0.2817
	M10	0.3048	0.7896	0.2051	0.5926	0.1938	0.5821	0.8063	0.2559
Sum Fusion	M1	0.2424	0.7587	0.1418	0.8333	0.2449	0.5656	0.7551	0.2779
	M2	0.0263	0.7861	0.0133	1	0.2145	0.5067	0.7855	0.1023
	M3	0	0.7399	0	0	0.2601	0.5	0.7399	0
	M4	0	0.7399	0	0	0.2601	0.5	0.7399	0
	M5	0	0.7399	0	0	0.2601	0.5	0.7399	0
	M6	0.3103	0.7683	0.1915	0.8182	0.2351	0.5878	0.7649	0.32
	M7	0.3478	0.7839	0.2564	0.5405	0.1871	0.5966	0.8129	0.2613
	M8	0.3768	0.7522	0.3333	0.4333	0.1812	0.6035	0.8188	0.2284
	M9	0.3238	0.7954	0.2179	0.6296	0.1906	0.5904	0.8094	0.2817
	M10	0.3048	0.7896	0.2051	0.5926	0.1938	0.5821	0.8063	0.2559
Weighted Average	M1	0.5164	0.8012	0.3901	0.7639	0.1928	0.6725	0.8072	0.4439
	M2	0.541	**0.8382**	0.44	0.7021	0.1405	0.6942	0.8595	0.467
	M3	0.3724	0.737	0.3	0.4909	0.2165	0.5651	0.7835	0.2287
	M4	0.4286	0.7688	0.3333	0.6	0.2027	0.6276	0.7973	0.3184
	M5	0.4167	0.7977	0.2778	0.8333	0.2057	0.6291	0.7943	0.4026
	M6	0.5887	0.8166	0.4823	0.7556	0.1706	0.712	0.8294	0.498
	M7	0.525	0.781	0.5385	0.5122	0.1358	0.6949	0.8642	0.383
	M8	0.3714	0.7464	0.3333	0.4194	0.1825	0.5998	0.8175	0.2174
	M9	0.56	0.8098	0.5385	0.5833	0.1309	0.7135	0.8691	0.4395
	M10	0.5235	0.7954	0.5	0.5493	0.1413	0.6902	0.8587	0.3943

Appendix A4: Results obtained on D4 (ReCOVery tweets)

Fusion	Model	F1	Accuracy	Precision	Recall	FPR	ROC	Spec	MCC
Early Fusion	M1	0.7939	0.8231	0.6818	0.9502	0.2481	0.823	0.7519	0.6735
	M2	0.7553	0.9172	0.6614	0.8803	0.0766	0.8199	0.9234	0.7171
	M3	0.7673	0.9243	0.6854	0.8714	0.0669	0.8314	0.9331	0.7303
	M4	0.7415	0.9222	0.6124	0.9397	0.0801	0.8018	0.9199	0.7203
	M5	0.7292	0.9202	0.5899	0.9545	0.0842	0.7918	0.9158	0.7127
	M6	0.8303	0.8517	0.7259	0.9698	0.219	0.8517	0.781	0.7267
	M7	0.756	**0.9273**	0.618	0.9735	0.0787	0.807111	0.9213	0.7413
	M8	0.7822	0.915	0.776	0.7884	0.0546	0.8625	0.9454	0.7294
	M9	0.7893	0.9191	0.7708	0.8087	0.0554	0.8631	0.9446	0.7397
	M10	0.7859	0.9191	0.7552	0.8192	0.0588	0.868	0.9413	0.7371
Average Fusion	M1	0.7776	0.8147	0.6477	0.9725	0.2641	0.8147	0.7359	0.6677
	M2	0.7485	0.914	0.7022	0.8013	0.0646	0.8317	0.9354	0.6991
	M3	0.7357	0.9007	0.7584	0.7143	0.0546	0.8454	0.9454	0.6751
	M4	0.7673	0.9243	0.6854	0.8714	0.0669	0.8314	0.9331	0.7303
	M5	0.7104	0.9007	0.6685	0.758	0.072	0.8105	0.928	0.6527
	M6	0.8223	0.8464	0.7111	0.9748	0.2274	0.8464	0.7726	0.7195
	M7	0.7538	0.9172	0.6561	0.8857	0.0776	0.8179	0.9224	0.7168
	M8	0.6981	0.869	0.7708	0.6379	0.0591	0.8319	0.9409	0.6199
	M9	0.7574	0.9161	0.6667	0.8767	0.077	0.8219	0.923	0.7175
	M10	0.744	0.912	0.651	0.8681	0.0804	0.8134	0.9196	0.7026
Max Fusion	M1	0.6916	0.7635	0.5303	0.9938	0.3203	0.7635	0.6797	0.5958
	M2	0.6716	0.9099	0.5056	1	0.0992	0.7528	0.9008	0.6749
	M3	0.674	0.9089	0.5169	0.9684	0.0975	0.7565	0.9025	0.6685
	M4	0.6716	0.9089	0.5112	0.9785	0.0984	0.7544	0.9016	0.6691
	M5	0.6543	0.9048	0.4944	0.967	0.1016	0.7453	0.8984	0.6516
	M6	0.7715	0.8136	0.6296	0.996	0.2708	0.8136	0.7292	0.6744
	M7	0.7516	0.9223	0.6085	0.9829	0.0859	0.803	0.9141	0.7372
	M8	0.7599	0.9191	0.651	0.9124	0.0798	0.8179	0.9202	0.7276
	M9	0.7524	0.9212	0.6094	0.9832	0.0874	0.8034	0.9126	0.7373
	M10	0.7403	0.9181	0.5938	0.9828	0.0906	0.7956	0.9094	0.7263
Sum Fusion	M1	0.6919	0.7639	0.5303	0.9953	0.3201	0.7639	0.6799	0.597
	M2	0.6716	0.9099	0.5056	1	0.0992	0.7528	0.9008	0.6749
	M3	0.674	0.9089	0.5169	0.9684	0.0975	0.7565	0.9025	0.6685
	M4	0.6716	0.9089	0.5112	0.9785	0.0984	0.7544	0.9016	0.6691
	M5	0.6543	0.9048	0.4944	0.967	0.1016	0.7453	0.8984	0.6516
	M6	0.7695	0.8099	0.6222	0.996	0.2747	0.8099	0.7253	0.6686
	M7	0.7516	0.9223	0.6085	0.9829	0.0859	0.803	0.9141	0.7372
	M8	0.7599	0.9191	0.651	0.9124	0.0798	0.8179	0.9202	0.7276
	M9	0.7524	0.9212	0.6094	0.9832	0.0874	0.8034	0.9126	0.7373
	M10	0.7403	0.9181	0.5938	0.9828	0.0906	0.7956	0.9094	0.7263
Weighted Average	M1	0.7939	0.8231	0.6818	0.9502	0.2481	0.823	0.7519	0.6735
	M2	0.756	**0.9273**	0.618	0.9735	0.0787	0.807111	0.9213	0.7413
	M3	0.7415	0.9222	0.6124	0.9397	0.0801	0.8018	0.9199	0.7203
	M4	0.7673	0.9243	0.6854	0.8714	0.0669	0.8314	0.9331	0.7303
	M5	0.7292	0.9202	0.5899	0.9545	0.0842	0.7918	0.9158	0.7127
	M6	0.8303	0.8517	0.7259	0.9698	0.219	0.8517	0.781	0.7267
	M7	0.7553	0.9172	0.6614	0.8803	0.0766	0.8199	0.9234	0.7171
	M8	0.7822	0.915	0.776	0.7884	0.0546	0.8625	0.9454	0.7294
	M9	0.7859	0.9191	0.7552	0.8192	0.0588	0.868	0.9413	0.7371
	M10	0.7893	0.9191	0.7708	0.8087	0.0554	0.8631	0.9446	0.7397

Appendix A5: Results obtained on D5 (CoAID)

Fusion	Model	F1	Accuracy	Precision	Recall	FPR	ROC	Spec	MCC
Early Fusion	M1	0.8601	0.8735	0.7925	0.9403	0.1737	0.9602	0.8263	0.7552
	M2	0.823	0.8009	0.9901	0.7042	0.0135	0.965	0.9865	0.657
	M3	0.776	0.8102	0.703	0.8659	0.2239	0.919	0.7761	0.6244
	M4	0.368	0.6343	0.2277	0.9583	0.4063	0.9421	0.5938	0.3477
	M5	0.84	0.8519	0.8317	0.8485	0.1453	0.9142	0.8547	0.7022
	M6	0.8845	0.8735	0.9874	0.801	0.0156	0.9846	0.9844	0.768
	M7	0.8791	0.8981	0.7921	0.9877	0.1556	0.9855	0.8444	0.8074
	M8	0.8731	0.8843	0.8515	0.8958	0.125	0.9532	0.875	0.7677
	M9	0.8479	0.8472	0.9109	0.7931	0.09	0.943	0.91	0.7026
	M10	0.91	0.912	0.9505	0.8727	0.0472	0.9715	0.9528	0.8272
Average Fusion	M1	0.9789	0.9785	0.9759	0.9818	0.025	0.9785	0.975	0.9569
	M2	0.8	0.8065	0.8155	0.785	0.1727	0.8069	0.8273	0.613
	M3	0.7727	0.7696	0.8252	0.7265	0.18	0.7723	0.82	0.5455
	M4	0.8455	0.8433	0.9029	0.7949	0.1	0.8462	0.9	0.6936
	M5	0.8	0.788	0.8932	0.7244	0.122	0.7931	0.8778	0.5941
	M6	0.9498	0.9538	0.9342	0.966	0.0562	0.9527	0.9438	0.9076
	M7	0.7834	0.7834	0.8947	0.6967	0.1053	0.7957	0.8947	0.5915
	M8	0.785	0.788	0.8842	0.7059	0.1122	0.7987	0.8878	0.5955
	M9	0.839	0.8479	0.9053	0.7818	0.0841	0.8543	0.9159	0.7031
	M10	0.8351	**0**.8579	0.8526	0.8182	0.1186	0.8525	0.8814	0.7023
Max Fusion	M1	0.8678	0.88	0.7711	0.9922	0.1939	0.8824	0.8061	0.7814
	M2	0.8817	0.8986	0.7961	0.988	0.1567	0.8937	0.8433	0.809
	M3	0.8681	0.8894	0.767	1	0.1739	0.8835	0.8261	0.796
	M4	0.9167	0.9263	0.8544	0.9888	0.1172	0.9228	0.8828	0.8585
	M5	0.914	0.9252	0.85	0.9884	0.1172	0.9228	0.8828	0.8561
	M6	0.7809	0.8308	0.6447	0.9899	0.2389	0.8195	0.7611	0.6927
	M7	0.8953	0.9171	0.8105	1	0.1286	0.9053	0.8714	0.8404
	M8	0.8706	0.8986	0.7789	0.9867	0.1479	0.8854	0.8521	0.804
	M9	0.908	0.9263	0.8316	1	0.1159	0.9158	0.8841	0.8574
	M10	0.869	0.8986	0.7684	1	0.1528	0.8842	0.8472	0.8069
Sum Fusion	M1	0.8639	0.8769	0.7651	0.9922	0.198	0.8794	0.802	0.7763
	M2	0.8817	0.8986	0.7961	0.988	0.1567	0.8937	0.8433	0.809
	M3	0.8681	0.8894	0.767	1	0.1739	0.8835	0.8261	0.796
	M4	0.9167	0.9263	0.8544	0.9888	0.1172	0.9228	0.8828	0.8585
	M5	0.914	0.9252	0.85	0.9884	0.1172	0.9228	0.8828	0.8561
	M6	0.7711	0.8246	0.6316	0.9897	0.2456	0.8129	0.7544	0.6824
	M7	0.8953	0.9171	0.8105	1	0.1286	0.9053	0.8714	0.8404
	M8	0.8706	0.8986	0.7789	0.9867	0.1479	0.8854	0.8521	0.804
	M9	0.908	0.9263	0.8316	1	0.1159	0.9158	0.8841	0.8574
	M10	0.869	0.8986	0.7684	1	0.1528	0.8842	0.8472	0.8069
Weighted Average	M1	0.9849	0.9846	0.9819	0.9879	0.0188	0.9847	0.9813	0.9692
	M2	0.9754	0.977	0.9612	0.99	0.0342	0.9762	0.9658	0.9541
	M3	0.9758	0.977	0.9806	0.9712	0.0177	0.9771	0.9823	0.9539
	M4	0.9854	0.9862	0.9806	0.9902	0.0174	0.9859	0.9826	0.9723
	M5	0.9709	0.9724	0.9709	0.9709	0.0263	0.9723	0.9737	0.9446
	M6	0.9699	0.9723	0.9539	0.9864	0.0393	0.9712	0.9607	0.9447
	M7	0.9405	0.9493	0.9158	0.9667	0.063	0.9456	0.937	0.8974
	M8	0.963	0.9677	0.9579	0.9681	0.0325	0.9667	0.9675	0.9344
	M9	0.9444	0.9539	0.8947	1	0.0758	0.9474	0.9242	0.9094
	M10	0.9399	0.9493	0.9053	0.9773	0.0698	0.9444	0.9302	0.8981

Appendix A6: Results obtained on D6 (MediaEval 2020)

Fusion	Model	F1	Accuracy	Precision	Recall	FPR	ROC	Spec	MCC
Early Fusion	M1	0.1429	0.8481	0.0851	0.4444	0.1401	0.6498	0.8599	0.1423
	M2	0.2264	0.8057	0.1935	0.2727	0.1323	0.5779	0.8677	0.1212
	M3	0.2326	0.8436	0.1613	0.4167	0.1307	0.5957	0.8693	0.1871
	M4	0.2667	0.8436	0.1935	0.4286	0.1269	0.6956	0.8731	0.2121
	M5	0.1081	0.8436	0.0645	0.3333	0.1415	0.5865	0.8585	0.0901
	M6	0.08	0.8544	0.0426	0.6667	0.1438	0.7021	0.8562	0.1425
	M7	0.0625	0.8578	0.0323	1	0.1429	0.5601	0.8571	0.1663
	M8	0.1176	0.8578	0.0645	0.6667	0.1394	0.6151	0.8606	0.1763
	M9	0.1765	0.8673	0.0968	1	0.1346	0.7566	0.8654	0.2894
	M10	0.1212	0.8626	0.0645	1	0.1388	0.5813	0.8612	0.2357
Average Fusion	M1	0.2857	0.8576	0.2045	0.4737	0.1178	0.5839	0.8822	0.2443
	M2	0.303	0.673	0.4545	0.2273	0.1241	0.584	0.8759	0.1316
	M3	0.4	0.8436	0.3333	0.5	0.1164	0.6358	0.8836	0.3227
	M4	0.2963	0.8199	0.2424	0.381	0.1316	0.5847	0.8684	0.2055
	M5	0.2	0.8483	0.1212	0.5714	0.1422	0.5522	0.8578	0.2117
	M6	0.2759	0.8671	0.1818	0.5714	0.1192	0.5799	0.8808	0.2688
	M7	0.303	0.673	0.4545	0.2273	0.1241	0.584	0.8759	0.1316
	M8	0.3137	0.8341	0.2424	0.4444	0.1295	0.5931	0.8705	0.2422
	M9	0.2917	0.8389	0.2121	0.4667	0.1327	0.5836	0.8673	0.2363
	M10	0.2381	0.8483	0.1515	0.5556	0.1386	0.5645	0.8614	0.232
Max Fusion	M1	0	0.8608	0	0	0.1392	0.5	0.8608	0
	M2	0.1111	0.8483	0.0606	0.6667	0.149	0.5275	0.851	0.1687
	M3	0	0.8436	0	0	0.1564	0.5	0.8436	0
	M4	0	0.8436	0	0	0.1564	0.5	0.8436	0
	M5	0	0.8436	0	0	0.1564	0.5	0.8436	0
	M6	0	0.8608	0	0	0.1392	0.5	0.8608	0
	M7	0.0571	0.8436	0.0303	0.5	0.1531	0.5123	0.8469	0.0925
	M8	0	0.8436	0	0	0.1564	0.5	0.8436	0
	M9	0	0.8436	0	0	0.1564	0.5	0.8436	0
	M10	0.0588	0.8483	0.0303	1	0.1524	0.5151	0.8476	0.1603
Sum Fusion	M1	0	0.8608	0	0	0.1392	0.5	0.8608	0
	M2	0.1111	0.8483	0.0606	0.6667	0.149	0.5275	0.851	0.1687
	M3	0	0.8436	0	0	0.1564	0.5	0.8436	0
	M4	0	0.8436	0	0	0.1564	0.5	0.8436	0
	M5	0	0.8436	0	0	0.1564	0.5	0.8436	0
	M6	0	0.8608	0	0	0.1392	0.5	0.8608	0
	M7	0.0571	0.8436	0.0303	0.5	0.1531	0.5123	0.8469	0.0925
	M8	0	0.8436	0	0	0.1564	0.5	0.8436	0
	M9	0	0.8436	0	0	0.1564	0.5	0.8436	0
	M10	0.0588	0.8483	0.0303	1	0.1524	0.5151	0.8476	0.1603
Weighted Average	M1	0.125	0.8671	0.0682	0.75	0.1314	0.5323	0.8686	0.1998
	M2	0.1111	0.8483	0.0606	0.6667	0.149	0.5275	0.851	0.1687
	M3	0.4151	0.8531	0.3333	0.55	0.1152	0.6414	0.8848	0.3506
	M4	0.2381	0.8483	0.1515	0.5556	0.1386	0.5645	0.8614	0.232
	M5	0.2	0.8483	0.1212	0.5714	0.1422	0.5522	0.8578	0.2117
	M6	0.0444	0.8639	0.0227	1	0.1365	0.5114	0.8635	0.1401
	M7	0.1579	0.8483	0.0909	0.6	0.1456	0.5398	0.8544	0.1903
	M8	0.2162	0.8626	0.1212	1	0.1401	0.5606	0.8599	0.3228
	M9	0.2381	0.8483	0.1515	0.5556	0.1386	0.5645	0.8614	0.232
	M10	0.2105	0.8578	0.1212	0.8	0.1408	0.5578	0.8592	0.276

End of the document.