IMPROVING FAKE NEWS DETECTION USING DEEP LEARNING TECHNIQUES

A THESIS

By

ROHIT KUMAR KALIYAR (Enroll. No.: E17SOE800)

BENNETT UNIVERSITY
TIMES OF INDIA GROUP

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

BENNETT UNIVERSITY

(Established under UP Act No 24, 2016)
Plot Nos 8-11, Tech Zone II,
Greater Noida-201310, Uttar Pradesh, India.

MAY, 2021

IMPROVING FAKE NEWS DETECTION USING DEEP LEARNING TECHNIQUES

Thesis submitted in fulfillment of the requirements for the award of the degree of

DOCTOR OF PHILOSOPHY

By

ROHIT KUMAR KALIYAR (Enroll. No.: E17SOE800)

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

BENNETT UNIVERSITY

(Established under UP Act No 24, 2016)
Plot Nos 8-11, Tech Zone II,
Greater Noida-201310, Uttar Pradesh, India.

MAY, 2021

DECLARATION BY THE SCHOLAR

I hereby declare that the work reported in the Ph.D. thesis entitled **"IMPROVING FAKE NEWS DETECTION USING DEEP LEARNING TECHNIQUES"** submitted at **Bennett University, Greater Noida, India** is an authentic record of my own work carried out under the supervision of **Dr. Anurag Goswami**, Assistant Professor, Department of Computer Science Engineering at Bennett University and **Dr. Pratik Narang**, Assistant Professor, Department of CSIS, BITS Pilani.

I have not submitted this work elsewhere for the award of any other degree or diploma. I am fully responsible for the contents of my Ph.D. Thesis.

ROHIT KUMAR KALIYAR
Department of Computer Science Engineering
Bennett University, Greater Noida, India
Date: May 9, 2021

SUPERVISOR'S CERTIFICATE

This is to certify that the work reported in the Ph.D. thesis entitled **"IMPROVING FAKE NEWS DETECTION USING DEEP LEARNING TECHNIQUES"** submitted by **Rohit Kumar Kaliyar** at **Bennett University, Greater Noida, India** is a bonafide record of his original work carried out under my supervision. This work has not been submitted elsewhere for any degree or diploma.

(Signature of Supervisor-1)
Dr. Anurag Goswami
Bennett University, Greater Noida, India
Date: May 9, 2021

(Signature of Supervisor-2)
Dr. Pratik Narang
BITS Pilani, Pilani, India
Date: May 9, 2021

Contents

List of Figures

List of Tables

Abbreviations

CNN- Convolutional Neural Network
LSTM- Long Short Term Memory
RNN- Recurrent Neural Network
ML- Machine Learning
DL- Deep Learning
DNN- Deep Neural Network
SVM- Support Vector Machine
DCNN- Deep Convolutional Neural Network
BERT- Bidirectional Encoder Representations from Transformers
RF- Random Forest
FPR- False Positive Rate
FNR- False Negative Rate
TFM- Tensor Factorization Method

Acknowledgment

First of all, I would like to acknowledge my mentors, Dr. Anurag Goswami and Dr. Pratik Narang, to introduce me to the world of research. It has been almost four years since then, and there has been no turning back. His passion for applied engineering and willingness to explore new things have helped me throughout this journey. It was a real privilege and an honour for me to share Dr. Anurag's exceptional scientific knowledge and his extraordinary human qualities and similarly for Dr. Pratik Narang. I have learned so many things from them, related to both the subject and life in general. I always enjoy discussions and conversations with them. Thank you, sir, for your constant support and guidance. It was a great pleasure for me to have a chance of working with them. I want to express my gratitude to the Research Advisory Committee (RAC), Prof. Deepak Garg (HOD CSE), Prof. Shivani Goel, and Prof. Vinayak Ranjan for constructive comments and guidance.

I want to thank all the Faculty of Computer Science Engineering (CSE) faculty and staff members at Bennett University (The Times Group) Greater Noida for their help and support. I am deeply thankful to Shri Vineet Jain, Chancellor, Prof. Prabhu Agrawal, Vice-Chancellor, Bennett University, Col. Guljit Singh Chadha (Retd.), Registrar, Prof. Rama Komaragiri, Dean (Academic), Prof. Debajyoti Mukhopadhyay (Dean, School of Engineering & Applied Sciences) and Prof. R. S. Chauhan, (Dean R & C), for their moral support and providing me all the necessary facilities during my candidature. I acknowledge Bennett University (The Times Group) for providing me with the scholarship to pursue doctoral studies during my candidature. I acknowledge my thanks to the Bennett University Library for allowing me to access valuable books and journals. I would also like to thank the administration of Bennett University, in which Mr. Pawan Kumar, Mr. Arvind Mishra were most appreciable for helping me always in an amicable manner.

I would also like to thank my parents (Mr. Yashpal Singh Kaliyar and Mrs. Sushila Devi) and my wife (Vishika) for their willingness to support me through higher studies. My special gratitude to Vishika, who is the best partner one can hope to have, for her love and support. It is beautiful to have someone willing to listen to all the technical jargon patiently. There have been times when I have gotten new ideas from our discussions. This journey would have been difficult without your constant presence in my life. A special thanks for her love, affection, continuous support and encouragement. Without my family members' complete devotion and support, there would have been nothing to write an acknowledgment.

I would like heartily thank my research colleagues for their never-ending moral support to make my PhD work easier. Thank you for constantly pushing and helping me to develop an outstanding work-life balance. I would also like to thank my colleagues in HCL, Noida, who inspired me to take this PhD. Program. Thank you for re-igniting the spark. I still maintain that theory is more important than applications. I would also like to mention my sincere gratitude towards my juniors for their help and support. I have been privileged to be surrounded by such curious people who have taught me never to forgo the eagerness to learn new things.

Subsequently, I would like to show my sincere appreciation for Bennett University, my Alma mater. The quiet, lush green campus, the historical and magnificent infrastructure, when I needed a long walk to clear my head. Furthermore, I cannot thank the well-maintained Bennett Central Library for providing access to almost all internationally acclaimed journals. Finally, I acknowledge everyone who has been a part of my life for helping me, in one way or the other, through the course of this endeavor.

Dedicated to

My Mentors, My Wife
&
My Parents

Abstract

Social media can provide instant news faster than conventional news outlets or sources. It is a great wealth of information, yet there is a growing need to verify this information's accuracy and correctness. The rate of producing digital information is large and quick, running daily at every second. The developing fame of social media has brought the massive creation of user-generated content. A significant part of this data is valuable and has turned out to be a great learning source. The growing fame of social media outlets has estimated the distribution of news articles that have caused the fake news explosion. Fake news is made and published with the intent to mislead and damage the representation of an agency, entity, person for commercial and political advantages. With the rapid emergence of fake news, serious attentiveness has produced in our society due to immense fake content distribution. The widespread of fake news has the potential for incredibly adverse effects on individuals and civilization. In this regard, fake news identification via social media platforms has, as of late, turned into an emerging research topic that is drawing huge consideration. Currently, there is no direct method to distinguish whether the information presented as a piece of news is both trustworthy and beneficial. Search engines are the doors to learning, but seeking significance cannot ensure that the matter is reliable. An easygoing observer probably won't have the capacity to differentiate between reliable and untrustworthy news. My research work is centered on evaluating such imparted news articles on social media for their reliability and trustworthiness. Fundamental theories of trust have utilized to motivate the search for a better solution.

Chapter 1

Introduction

Social media platforms, along with the change in the mobile technology, has gained popularity by reaching on the fingertips of the users. In the past few years, social media platforms such as Twitter, Facebook, Instagram, etc., also became very influential since they facilitate the smooth acquisition of information and provide a quick outlet for sharing false information (Kumar & Shah 2018). This false information is often called as misinformation or disinformation. In society, misinformation can be created and shared quickly by using web-based social networking platforms., bringing about an overall real-world impact on individual and society. However, disinformation is especially hazardous because it is organized, well resourced, and reinforced by automated technology (Kumar & Shah 2018). A recent survey (Zhou & Zafarani 2018) has alarmingly demonstrated that people progressively obtain the daily news from web-based social platforms rather than traditional news sources, making it quite significant to reduce false data toward such stages. With primary purposes of changing opinions and making money (Kumar & Shah 2018; Zhou & Zafarani 2018), the broad impact of inaccurate knowledge makes it one of the advanced threats to civilization, as designated by the World Economic Forum (Zhou & Zafarani 2018).

Defining how misinformation multiplies on social stages and why it prevails regarding misleading readers are essential ideas to create a useful detection model or early detection of inaccurate content. An ongoing research surge here has meant addressing the key issues and utilizing feature engineering (Kumar & Shah 2018), graph mining (Zhou & Zafarani 2018), and information modelling (Kumar & Shah 2018). The more significant part of the research has principally centered around two general classifications of false data: sentiment-based and truth-based/fact-based. It is difficult for naive users with some electronic data to examine its dependability or reliability.

The availability of unauthentic data on social media platforms has attracted researchers' attention and become a hot-spot for detecting fake news (Vosoughi et al. 2017; Zubiaga et al. 2016) effectively. Nowadays, fake news is treated as an important issue due to its negative impact (Zhou & Zafarani 2018; Zubiaga et al. 2016) and has gained attention among researchers, journalists, politicians and the general public. In the context of writing style, fake news is communicated or published to mislead the people and damage the representation of an agency, entity, person, either for commercial or political advantages (Ghosh & Shah 2018; Ruchansky et al. 2017; Zhou & Zafarani 2018). Few examples of fake news are shown in figure 1.1. These examples of fake news were trending throughout the COVID-19 pandemic and the 2016 U.S. General Presidential Election. In the research context, relevant synonyms (keywords) often linked with fake news (refer to Table 1.1):

FIGURE 1.1: Examples of Fake News during 2016 U.S. General Elections and during the pandemic of COVID-19 (Source: Facebook and Twitter)

- Rumor: A rumour (Bondielli & Marcelloni 2019; Fazil & Abulaish 2018; Gorrell et al. 2018) is an unverified claim about any event, transmitting from individual to individual in society. It might signify an occurrence, article, and any social issue of genuine public concern. It might end up being a socially threatening phenomenon in any human culture.
- Hoax: A hoax is a story deliberately made to masquerade as the truth (Tacchini et al. 2017). Currently, it has been growing at an alarming rate. Hoax is also known with similar names like prank or jape.
- Misinformation: Misinformation (Antoniadis et al. 2015; Kumar & Geethakumari 2014) is used to share incorrect information disregarding the actual intent. This information consist of false output labels.
- Disinformation: Disinformation (Shu et al. 2017) is used sharing a wrong intent to mislead society. Disinformation is disseminated tactically with some biased information and manipulated facts. It usually defined with the term "Propaganda".

TABLE 1.1: Concepts of fake news in social media

Concept	**Authenticity**	**Intention**	**News?**
Fake news	Non-factual	Undefined	Yes
Hoax	Non-unified	Entertain	Yes
Rumor	Undefined	Undefined	Not defined
Misinformation	Non-factual	Undefined	Not defined
Disinformation	Non-factual	Mislead	Not defined
Click-bait	Undefined	Mislead	Not defined

Generally, fake news appears as a piece of unauthentic information that quickly shares social media to mislead society. It is evident that due to sharing an expanded volume of information everyday (Ahmed et al. 2017), the quality of content suffers the truth's ground (refer to figure 1.2). The intention of spreading false information is to manipulate public opinion for commercial and political earnings (Brummette et al. 2018). Fake news has also shown adverse impacts on stock-prices and notable infrastructure investments (H. De Vreese 2001). One such example is about a bomb explosion news (Bozarth et al. 2020) in which former U.S. President Barack Obama got injured. This news annihilated 130 billion US dollars in the stock market within a few minutes, shows an immediate impact of fake news.

The main contributors to fake news are Fakesters. "Fakesters" advertise fake news with particular plans to deceive people and publish fabricated articles by distorting the fact behind it (Zhou & Zafarani 2019). One of the stakeholders fighting against the fake news are fact-checking organizations like Snopes [*] and Politifact [†] etc. These organizations are valuable in terms of validating the news content with facts-based methods. However, for checking the quality of news content, these methods are not automated and tend to be very time-consuming (Zhou & Zafarani 2019). It is also complicated to measure the quality of content daily by these methods (Kumar et al. 2020; Zhou & Zafarani 2019) for news created by different Fakesters. The complete life-cycle of fake news starting from creation till propagation has been described by (Kumar & Shah 2018). In this life cycle, the role of individuals (user, creator, publisher, etc.) has been examined with the help of a fake news article's overall structure. We can categorize the stages of fake news into three forms: creation (how a piece of fake news created), publication (publication and feedback of a fake article), and propagation (how a piece of fake news distributed among users in social media). Different characteristics are shown in figure 1.3.

[*] https://www.snopes.com
[†] https://www.politifact.com

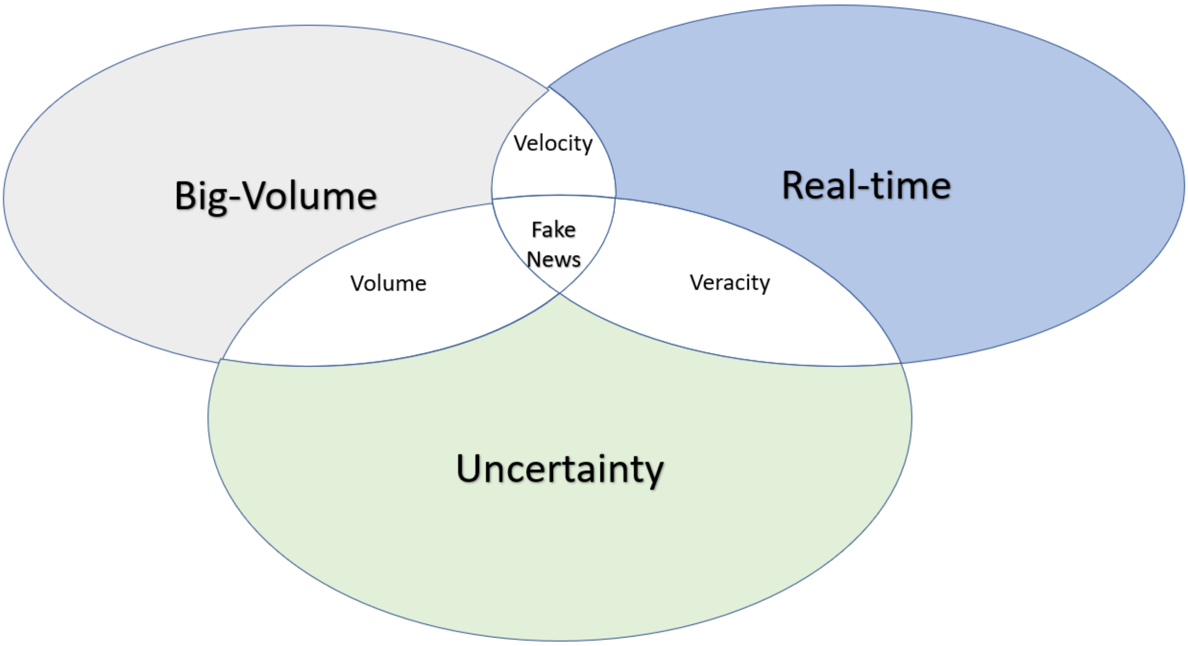

FIGURE 1.2: Fake News in Terms of 3 V

1.0.1 Why fake news detection is so important?

Fake news paves the way for misleading the people of society and developing ideologies. These people produce incorrect knowledge to others for some personal interest or making money with the number of inter-communications on their false statements (Brummette et al. 2018; Fourney et al. 2017). Disseminating disinformation endures numerous intentions, i.e., to gain support in political elections, business and product development, and personal retaliation (Brummette et al. 2018). Individuals can be susceptible to fake news, questioning to differentiate from the specific and accurate information. People in society are easily manipulated, especially by sharing the news via social media with their colleagues and family members due to relations and faith. We lead our sentiments on the news, making it satisfactory, when appropriate and starring from our own opinions. Therefore, we get convinced that it outlines and falls into the traps and shares it with the world. Thus, an effective fake news detection approach is a primary necessity in current situations.

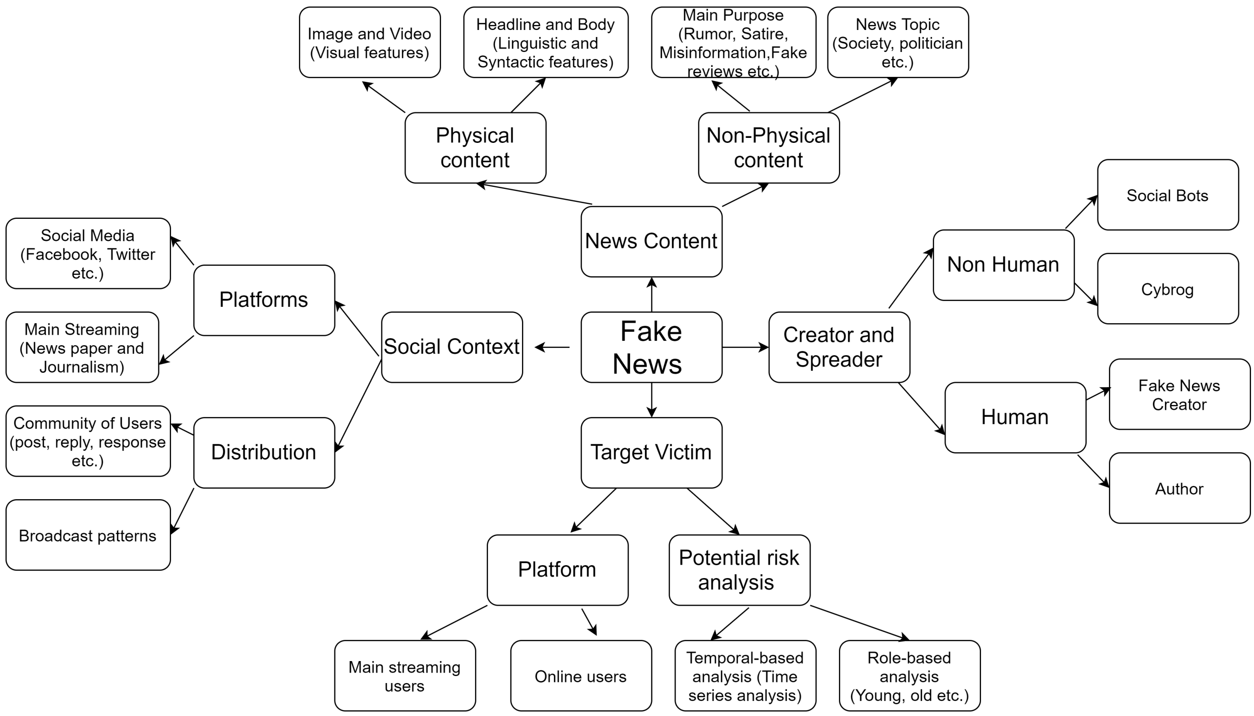

FIGURE 1.3: Characteristics of Fake News

1.0.2 Fundamental theories

In this section, fundamental theories for fake news detection have discussed. Different fact-checking methods, fact-checking sites have also been discussed.

- Knowledge-based Fake News Analysis (Ghosh & Shah 2018; Kumar & Shah 2018; Zhou & Zafarani 2018): These theories investigate the fake news detection using fact-checking which can be further categorized as follows:
  - Manual fact checking: It mainly aims to evaluate the authenticity of the news utilizing the knowledge obtained from verified information. It can be divided into two parts as below:
    * Fact checking by experts: It depends on the knowledge of domain experts to verify the news content available on social media.
    * Fact-checking by crowd-sourced : It depends on a large society of people leading as valuable fact-checkers.

  - Automated fact-checking: This process depends upon the approaches used to solve Natural Language Processing (NLP) tasks and graph theory-based problems. It can be divided into three parts as below:
    - ∗ Fact extraction: Knowledge is extracted from different web sources.
    - ∗ Knowledge processing tasks: It includes redundant facts, conflicting facts and unreliable knowledge.
    - ∗ Fact-checking: It is a process to estimate the authenticity of news articles available on social media.

- Style-based Fake News Analysis (Ahmed et al. 2017; Kumar & Shah 2018; Zhou & Zafarani 2018): These basic theories explained the role of fake news content and its writing style, which can differ from an accurate news article. Style-based fake news can be further categorized as follows:
  - Deception analysis and identification: It investigates deceptive content style across numerous types of information available on social media.
    - ∗ Deception style theories: Its main aim is to investigate various styles of false information that aim to deceive readers.
  - Style-based feature and patterns: It mainly explores the different style of news articles and the news pattern residing in them. Before this research, the progress of style-based fake news investigations was in its primary stages, and an inadequate number of studies were available. It can be categorized into two parts as below:
    - ∗ Attribute-based language features: These are inspired by news articles or directly derived from related deception as mentioned in earlier theories.
    - ∗ Structure-based language features: These features describe the different news content styles in four language levels: lexicon, syntax, semantic, and discourse.
  - Deception detection strategies: Its aim to investigate the style of deceptive content.

- Propagation-based Fake new Analysis (Kumar & Shah 2018): These theories investigate empirical patterns and categorization & comparison techniques that utilize such designs, models, or detection techniques. It can further categorize propagation-based fake news as follows:
  - Fake news propagation patterns: It only investigates the propagation patterns present in fake news and compares fake news propagation to regular news communication.

  - Models for fake news propagation: It only describes the propagation dynamics and detection models related to fake news.
  - Propagation-based fake news detection (cascade based): It only describes the related patterns and models that categorize fake news propagation.
    * Utilizing cascade-similarity: To estimate the similarity between different types of news articles containing distinct domains and knowledge.
    * Utilizing cascade representation: It describes the different representation available in the given news information that can be utilized as features.

- Fake news detection using propagation dynamics (network-based):
  - Homogeneous Network: Consist of a single type of node and edge.
  - Heterogeneous Network: Consist of multiple types of nodes and edges.
  - Hierarchical Network: It consists of numerous types of nodes and edges present within data in the form of relationships.

- Credibility-based Fake News Analysis (Kumar & Shah 2018; Zhou & Zafarani 2018): These approaches examine the user's viewpoint for the effective detection of fake news. It further explored the relationship among news, user, and the user roles that a user can play in fake news creation and propagation. Credibility-based fake news can be further categorized as follows:
  - Assessing News Headline credibility: It often decreases to identifying click-bait's in social media data
  - Assessing News Source Credibility: To investigate the credibility of fake news stories circulated directly from social media websites. These websites only publish hoax-based news articles that are different from real news.
  - Assessing News comments credibility: It can be further divided into different detection models as follows:
    * Content-based methods: It estimates the credibility of a user's comment using a series of language features extracted from these comments.
    * User's behavior-based models: These models leverage distinct features of misleading comments obtained from metadata linked with user behaviors present in news articles.

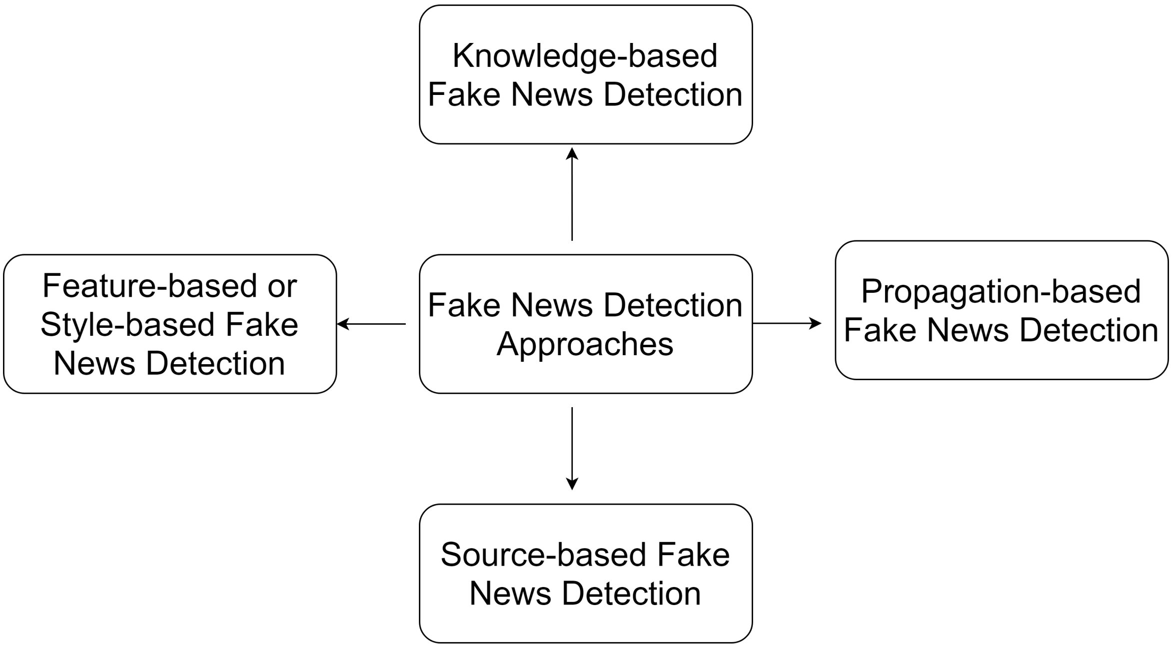

FIGURE 1.4: Approaches for Fake News detection

  * Graph-based models. It describes the relationships among reviewers, comments, products, etc.
- Accessing News Spreader credibility: Online users can be grouped into two types:
  * Naive users: These normal users unintentionally engage in fake news propagation.
  * Malicious users: They deliberately circulated fake news on social media.

1.0.3 Approaches for fake news detection

Detection of fake news is questionable because news is intentionally written to falsify accurate and authentic information. Fundamental theories, as mentioned earlier, were valuable in supervising further research in the field of fake news classification. Current methods for the detection of fake news can be categorized as:

- Knowledge-based detection methods: These methods mainly deals with the process of fact-checking. Fact-checking procedures primarily use news authenticity to improve journalism. This section will discuss manual fact-checking approaches and incorporate them toward automatic detection of fake news. These fact-checking methods can be divided as follows:

- Expert-based manual fact-checking: Fact-checkers handle this fact-checking process to verify the news content. It accompanies a small society of highly credible fact-checkers. These methods provide effective results to detect fake news, but this process is too costly and time-consuming when dealing with a high volume of news content.
- Expert-based fact-checking websites: Recently, numerous websites developed, allowing expert-based fact-checking for any news article circulated on social media. Some of these types of websites are:
  - PolitiFact [‡]: Mainly topics covered are related to American politics. Evaluations labels are True, Mostly True, Half true etc.
  - Snopes [§]: Mainly topics covered are politics, social, and topical issues. Evaluations labels are True, Mostly True, False, Scam, Legend, Outdated etc.
  - FactCheck [¶]: Mainly topics covered are related to American politics. Evaluations labels are True, No evidence, and False.
  - GossipCop [‖]: Mainly topics covered are related to Hollywood and celebrities. Evaluations labels are 0-10 scale, 0 shows completely fake news, and 10 indicate entirely trustworthy news.
  - TruthOrFicton [**]: Mainly topics covered are related to politics, religion, nature, food, and medical. Evaluations labels are truth or fiction.
  - FullFact [††]: Mainly topics covered are related to economy, health, crime, law, and immigration. Evaluations of labels are ambiguous. PolitiFact and GossipCop were the primary resource for the development of real-world fake news dataset: (e.g.LIAR (Wang 2017), and FakeNewsNet (Shu et al. 2018)).
- Crowd-sourced manual fact-checking: It depends on a large community of people working as fact-checkers. The group of such credible fact-checkers can gather from some reputed organizations or online business communities. It is significantly less credible, challenging to manage, and politically biased as compared to expert-based fact-checking.

[‡] https://www.politifact.com/
[§] https://www.snopes.com/
[¶] https://www.factcheck.org/
[‖] https://www.gossipcop.com
[**] https://www.truthorfiction.com/
[††] https://fullfact.org/

- Crowd-sourced fact-checking websites: In recent time, several websites emerged for allowing crowd-based fact-checking. Some of the websites are:
  * Fiskkit [‡‡]: Using this website, the user can upload articles and keep the tag that best describes the news articles. It is helpful to distinguish between news and non-news articles. Not so many crowd-sourcing websites exist, so it remains an open challenge for organizations like Facebook and Twitter to develop such tools for improving the quality of news content.
- Automated fact-checking: These methods are dependent on Machine Learning (ML) classification models and the techniques used for Natural Language Processing (NLP) tasks. It can be divided into two stages which are as follows:
  * Fact Extraction: Such a process is known as knowledge extraction or relation extraction from to-be-verified news articles on social media. The main issue like redundancy, invalidity, conflicts, and unreliability of data, will be handled in this process. Fact extraction methods have discussed in the research article (Rashkin et al. 2017).
  * Fact-checking: In this process, we need to compare the extracted knowledge with the existing facts. The main concern in these types of methods is the sources from which facts are extracted. Such references have rarely discussed in current research. However, fact-checking methods must identify the part-to-be-verified news is check worthy or not..

• Feature-based or Style-based detection methods: Style-based methods mainly handle the intention lies in the news, i.e., is there any intention in a news article to mislead the public or not?. The end goal of style-based methods is to capture a unique writing style in different malicious news articles used to mislead ordinary people. The classification performance of style-based detection methods dependent on the following points: effectively capture the writing style in news content and the performance of classifier based on the different news content representations. Generally, content-based features can be classified into textual features and visual features (Qassim et al. 2018), representing news text and images. In this research, news text was the primary resource. Textual features can be further categorized into two forms:

[‡‡] `https://fiskkit.com/`

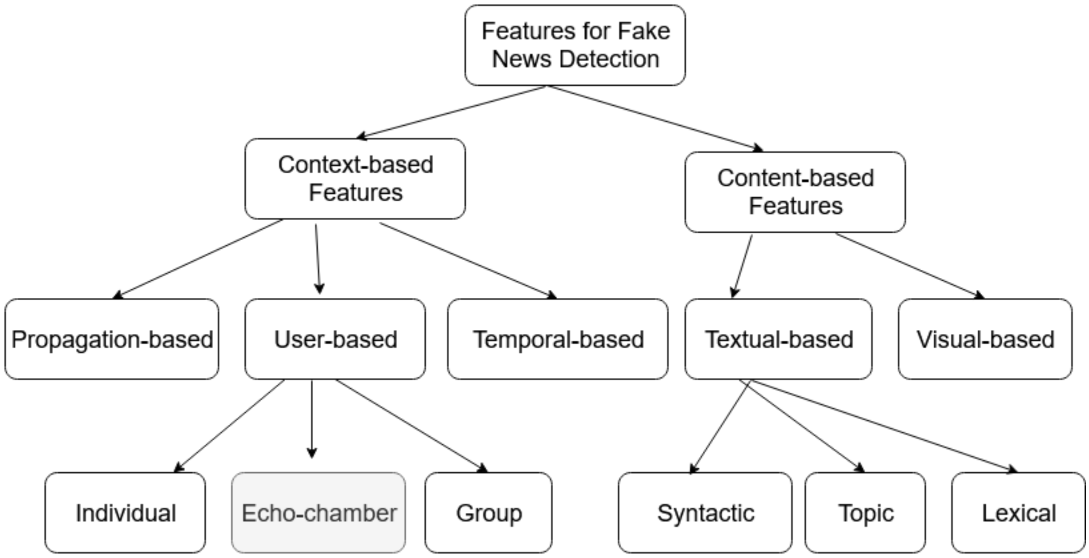

FIGURE 1.5: Features for Fake News detection

- General textual features: These features primarily used to detect fake news utilizing traditional ML frameworks. These features represent the content style from three language levels: lexicon, semantic, and syntax. The lexicon level's main task is to estimate the frequency of lexicons using a Bag-Of-Word (BOW) approach (Zhou & Zafarani 2018). Syntax level tasks performed with the help of Part-of-Speech (POS) taggers (Feng et al. 2012; Zhou & Zafarani 2019). At the semantic level, these processes can be assigned to lexicons-based analysis with a psycho-linguistic approach named LIWC: Linguistic Inquiry and Word Count (Pérez-Rosas et al. 2017).
- Latent textual features: These features primarily used with a combination of news-text embedding as word-level (Mikolov et al. 2013b), sentence-level (Mikolov et al. 2013b) or document level; results are vectors describing news articles and directly used as an input to the classifiers. These embedding can be further incorporated with neural architectures (Wang et al. 2016).
- Check the quality of content: To examine the quality of content for fake news detection, we require to extract valuable features (refer figure 1.5 for more details) from social media news datasets (Ahmed et al. 2017; Kumar & Shah 2018; Shin et al. 2018). Only a few datasets have all relevant information and features; most of them contain only linguistic features. Few datasets contain semantic and social contexts-based features. Nowadays, news fabrication is mostly occurring with textual content.

Existing techniques for the detection of fake news can be further categorized as follows:

- News content-based learning or News representation learning: News content-based methods (Ahmed et al. 2017; Ghosh & Shah 2018; Kumar & Shah 2018; Zhou & Zafarani 2018) primarily focus on extracting numerous features from fake news articles, including both content as well as style-based. Furthermore, fake news publishers commonly produce critical plans to spread damaged news, requiring specific writing styles (Zhou & Zafarani 2018) to interest and convince a wide extent of consumers who are not present in true news stories. Knowledge-based methods intend to use external sources to fact-check (Kumar & Shah 2018) news content's claims or truthfulness. Linguistic features like news content can be applied to find clues between real and fake information. Thus, it was challenging to identify fake articles more accurately by using only news content-based features (Shu et al. 2019d; Zhou et al. 2020). Consequently, there was a primary need to investigate the engagement of fake news articles with different users available online. Information-based methodologies (Roy et al. 2018; Zhou & Zafarani 2019) intend to utilize external sources to check news content claims' worth. It was evident that for efficient fake news detection, content-based methodologies were alone not sufficient. An investigation fake news articles with social context-based methods was the main necessity.
- Social context-based learning or Social engagement learning (Long et al. 2017; Ruchansky et al. 2017; Shu et al. 2017): In the current social computing era, social-engagements is another important feature for fake news detection. Social context-based approaches utilize user's social engagements as supporting knowledge to identify fake news effectively. Social context-based methodologies (Ruchansky et al. 2017; Sharma et al. 2019) deals with the relationship among users, news article, and publishers. These methodologies are valuable in recognizing fake news articles with high accuracy. Features related to fake news detection are shown in figure 1.5. Social engagements (the semantic relationship between news articles and user) can be vital for fake news detection. Approaches related to fake news identification showed in figure 1.4. Most of the existing and useful methods (Ruchansky et al. 2017; Sharma et al. 2019) consisted of news content and context level feature using unidirectional pre-trained word embedding models (such as GloVe, TF-IDF, word2Vec etc.) There was an enormous scope to use bidirectional pre-trained word embedding models having powerful feature extraction capability. In the current era

of computing, at any social media platform, a user was always connected to a specific group of peoples having the same mindset or liking is called a user-community. These user communities can be crucial for fake news classification due to their common perception about sharing articles. They are a group of users with the same interests in social circles where conflicting ideas are rejected and disapproved by the majority. For example, when looking at comments to a user post on Facebook, many of the comments that agree with the post reflect one echo chamber type. This understanding, combined with the others (e.g., number of likes, shares), may lead the user to get a false theory. Moreover, by default, Facebook increases the ranking (Shu et al. 2018, 2019d) of the comments based on the number of replies and likes received by the user's friends. Based on the issues discussed above, to design an efficient deep learning model was the primary motive utilizing social context-based features.

- Propagation-based detection methods: Propagation-based approaches (how a piece of news is shared among users using diverse context-related social media platforms) mainly deal with disseminating fake news. It also uses semantic relationships among social media posts and credibility scores (Shu et al. 2019b) by propagating preferences between users, user comments, and news articles. Instance-based methodologies (Sharma et al. 2019; Shu et al. 2019b) help to know users' perspectives from a sharing news article to induce unique news stories' integrity. Furthermore, propagation-based methodologies (Sharma et al. 2019) were based on the relations of valuable posts in social media to guide the learning of validity scores by obtaining credibility values (Shu et al. 2019b) between users, posts, and news. Propagation-based methods can be categorized into two types of process:

  – Fake News identification using news cascades: These methods adopted by the authors (Castillo et al. 2011; Ma et al. 2016; Vosoughi et al. 2017; Wu & Gu 2015) in their investigation. A news article can manage multiple simultaneously cascades due to different type of users available on social media.

  – Detection of fake articles using self-defined graphs: In these techniques, exploration is based on a news article's propagation-graph (Ma et al. 2016).

  Propagation-based approaches are most robust to capture the writing style of different type of users. These methods are not useful for early fake news detection. It performs well when limited information is available.

- Source-based detection methods: Fake news can be identified effectively by evaluating the credibility of its sharing source. The credibility of news is dependent on the quality of the news content (Castillo et al. 2011). Different roles of news authors, publishers, and presented users explored in the research articles (Shu et al. 2018; Zhang et al. 2020; Zhou et al. 2020). The source-based methods can be further divided into two forms:
  - The credibility of news authors and publishers: It uses the recent studies, methods which can be further divided as follows:
    * Reliability of authors and publishers: Existing research showed that news authors and publishers exhibit homogeneity in any network. Each node represents a news-author relationship in a news author's network, which ushered in FakeNewsNet (Shu et al. 2018).
    * Web Spam detection: News journalists often advertise fake or real news on their websites. Identifying deceptive publishers is directly depend on low-credible websites. Different web ranking algorithms are available to examine the low credible websites (Han et al. 2012).
    * Resources for trustworthy publishers: Several resources can help to check the credibility of news publishers. Major resources are fact-check website [§§] and NewsGuard [¶¶], which provides expert-based evaluations and a rating of news articles.
  - To access the credibility of users: In social media, different types of users are available with different credibility score. These different types of users are:
    * Identifying Malicious Users: Some bots are created to deceive social media users. These methods discussed by (Ferrara et al. 2016). Millions of social bots were used in online discussion during the 2016 U.S. general presidential election. The bot detection model using deep learning techniques utilizing user-posts and behavior has been discussed by (Cai et al. 2017).
    * Identifying Normal Users: Fake reviews can attract both non-credible or malicious users and normal users (Shu et al. 2017). Normal users often spread false information on the web. The credibility of fraudulent users is low as compared to regular users. These methods have investigated by (Ferrara et al. 2016).

[§§] https://mediabaisfactcheck.com/
[¶¶] https://newsguard.com/

1.0.4 Research directions

The detection of fake news gained enormous attention and became a prime area of focus hotspot for the research community. With a detailed analysis of existing methods, research directions can be categorized mainly (Ahmed et al. 2017; Kumar & Shah 2018; Liu & Wu 2018; Roy et al. 2018; Zhou & Zafarani 2018) into five perspectives:

- Data-oriented: It examines different viewpoints of fake news data, such as data collection, psychological validation of fake news, and early detection of fake news.
- Feature-oriented: It primarily intends to investigate valuable features for identifying fake news from multiple data sources, such as news content and social context etc.
- Model-oriented: It penetrates the gateway toward building more relevant and effective models for fake news detection, including different learning techniques: supervised, semi-supervised and unsupervised approaches.
- Application-oriented: It mainly includes the investigation approaches beyond the fake news detection, such as fake new diffusion and intervention.
- Cross-domain-oriented: It essentially investigates fake articles beyond domains, topics, and languages to obtain a knowledge of different news patterns and their features.

1.0.5 Industrial applications where the detection of fake news is useful

With a detailed analysis, research or industrial applications toward fake news detection as follows:

- Better fact-checking tools (Zhou & Zafarani 2018): Towards the designing of valuable online social platforms for fact-checking digital media to reject or authenticate disinformation.
- Public alert systems (Kumar & Shah 2018): To provide the data on significant disinformation campaigns to the public in real-time. It would provide researchers and the public an increased awareness of the activity and the ability to assess it.
- Bolstering Journalism (Kumar & Shah 2018): Government can assemble consistent professionals and encourages the development of partnerships that ensure reliable financing.

- Media literacy surge (Rashkin et al. 2017; Shu et al. 2017): To educate the public using best practices in evaluating the quality of information online.
- Government policies and National security (Castillo et al. 2011; Rashkin et al. 2017): It can help make security policies when the sources have re-hidden content is often false or slanted.
- A Helping tool for social media platforms (Kumar & Shah 2018): Towards an effective and accurate detection of online misinformation available on numerous social media platforms.
- A Helping tool for search engines for better classification performance (Shu et al. 2017).

1.1 Motivation and research goal

Fake news detection has gained massive attention from researchers across the world. Social media platforms have become a famous bridge among users (Fazil & Abulaish 2018; Vosoughi et al. 2017) for quick and seamless access to fake news. More evident than during the 2016 United States General Presidential Elections (Zhou & Zafarani 2018), Demonetization in India, and during the pandemic of COVID-19.
Due to the availability of numerous social media platforms, it is quite trendy to create fake news and share it with the world quickly. Existing detection methods (Fazil & Abulaish 2018; Feng et al. 2012; Ghosh & Shah 2018; Gupta et al. 2018; Pérez-Rosas et al. 2017; Ruchansky et al. 2017; Shu et al. 2019b, 2018, 2017) primarily focused on either content or social context-based information extracted from news articles. Fake News detection was still in its infancy with low accuracy using different real-time datasets with various features. Despite experiencing massive attention from the leading research communities worldwide, we were motivated to design an effective detection model to examine news content, social context, and community-level features with a tensor-factorization method. Existing detection methods mainly focused on two components:

- Valuable Features: Features that can help investigate the deceptive writing style across numerous topics and languages.
- Efficiency of the classification model: To increase classification accuracy by utilizing a small/large amount of available information in the form of news articles.

1.1.1 Research Questions

The above challenges created many research questions and demanded solid solutions. This research work was motivated by the research question mentioned below:
RQ1: Can we build a more accurate system or model for fake news detection?
RQ2: Would the combination of all content-context-user level features improve the classification performance?
RQ3: Which classification approach (machine learning or deep learning) is the most accurate for fake news detection using several real-world fake news datasets?
RQ4: Can we build a generalized model for fake news detection irrespective of dataset selection?

1.1.2 Promise of Deep Learning

Deep learning methods have an excellent prospect of fake news detection, utilizing multiple features. Deep learning methods are well-recognized for producing satisfactory results in a broad spectrum of artificial intelligence-based problems. Deep learning techniques successfully placed as replacement models into existing natural language systems that achieved better classification performance. Some of the motivations/promises behind this research were:

The Promise of Feature Learning

Deep learning methods can automatically obtain the required features from the natural language, rather than extracting the required and specified features by a domain expert.

The Promise of Continued Improvement

The performance of deep learning models to solve different natural language processing-based problems was excellent to achieve more accurate results. The improvements appear to be continuing to speed up the classification performance.

The development of End-to-End Models

To develop a large end-to-end deep learning model that can be fitted in different natural language problems suggesting a more general and better-performing approach.

Concretely, different architectures of neural networks have used to detect fake news and yielded remarkable results. This research attempted to utilize the power of deep learning to improve the classification outcome of existing fake news identification/detection systems.

Research Goal: *Improving Fake News Detection using Deep Learning Techniques*

Chapter 2

Related work

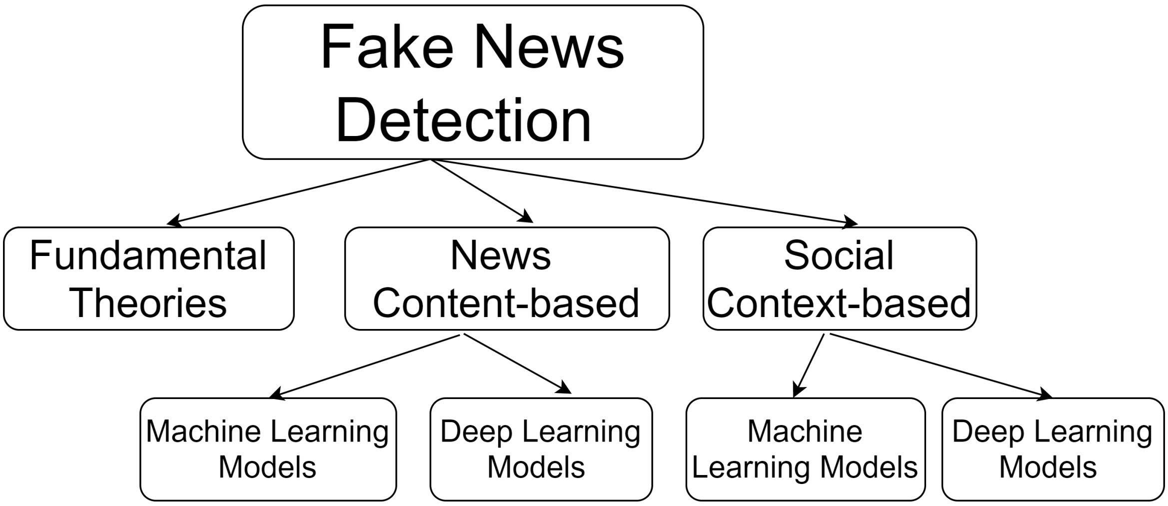

FIGURE 2.1: Existing Approaches for Fake News Detection

Despite the rapid increase in popularity, fake news detection is still in its new among researchers. There has been an increase in the number of examined methods (traditional machine learning and deep learning models) focusing on news content to detect fake news and rumours (refer to figure 2.1). This section represented real-world fake news datasets, current research work, research gaps identified, and research methodology used in our investigation.

2.1 Datasets

News articles can be identified as fake or real based on author or publisher, headline, text and news content. In the existing literature for fake news detection, various data-sets that were utilized are tabulated in table 2.1.

2.2 Fundamental theories

Current section explored the existing fundamental theories in the area of fake news identification.

2.2.1 An overview of fake news available online its impact on social media

In current section, a comprehensive survey on the impact of online fake news, existing detection methods, diverse aspects of fake news, and types of misleading information available on social

TABLE 2.1: Datasets in the Existing Literature

Dataset	Content	Number of Instances	Output Label
LIAR (Wang 2017)	Political statements	12800	Six labels (true, predominantly true, half-true, almost true, false, pants-fire)
Kaggle Fake News Dataset *	News during 2016 U.S General Election	20800	Binary (true or false)
CREDBANK (Mitra & Gilbert 2015)	Social network posts (Twitter)	37 million	Vector with 30 dimensions containing variable scores at five levels of veracity
FA-KES (Salem et al. 2019)	Events around Syrian war	804	Real or Fake
FAKENEWSNET (Shu et al. 2018)	News Articles	23921	Binary (true or false)
BuzzfeedNews (Shu et al. 2018)	Social network posts (Facebook)	2282	Four output labels
Emergent (Ferrara et al. 2016)	Related statements and titles	300	Binary (true or false)
Buzzface (Santia & Williams 2018)	Social network posts (Facebook)	2263	Four output labels (predominantly true, predominantly false, mix of true and false and no factual content)
Fake.Br Corpus (Monteiro et al. 2018)	News Articles	7200	Binary (true or false)
BuzzFeed-Webis (Potthast et al. 2017)	Social network posts (Facebook)	1687	Four output labels
CoAID (Cui & Lee 2020)	News Articles during COVID-19	4251	Binary (true or false)

media presented. In one study, Kumar et al. (Kumar & Shah 2018) investigated a broad survey of diverse aspects of fake news available on social media platforms. Different categories of fake news (e.g., misinformation, disinformation, rumours, hoax etc.), existing detection methods, and future aspects examined in their investigation propagation path and different search patterns. This research article's main objective was to present many types of misleading information available online and existing detection methods. In one of the research, Shin et al. (Shin et al. 2016) explored fundamental theories beyond different disciplines to magnify the interdisciplinary study of fake news. Sharma et al. (Sharma et al. 2019) presented the problems related to fake news and their technical challenges like early detection and combining different features associated with it. The authors discussed the existing and relevant methods to identify the propagation path, improving in each process and their limitations. To alleviate the quality of existing datasets, they comprehensively organized and summarized characteristic features of available datasets. Advances in each method, disadvantages, datasets, and limitations have also discussed. Zhang et al. (Zhang & Ghorbani 2020) presented a comprehensive survey on the existing detection methods with the social impact on society. The authors have also discussed datasets, different types of features, and future research directions in fake news. Technical challenges like early detection of fake news and tracing fake news instances over some time also discussed. They were using the real-world dataset to detect fake news effectively, also discussed in their research article.

In one investigation, Su et al. (Su et al. 2020) presented an overview of other misinformation available online at different social media platforms and their issues and detection approach. The authors presented their research in terms of detection methods, feature representation, evaluation matrices, and datasets. Advantages (less complexity, fast learning to classification model) and disadvantages (more features needed for effective detection, diverse nature of fake news) also discussed, focusing on content-based analysis. Batrinca et al. (Batrinca & Treleaven 2015) presented a survey of software tools for social networking. They have also introduced a study of different and valuable methodologies and critics of social media tools for fake news propagation. Fernandez et al. (Fernandez & Alani 2018) presented a survey of detection methodologies, research applications, and available online misinformation limitations (authenticity, news or not etc.). The authors also discussed existing fake news classification methods. Zhou et al. (Zhou et al. 2020) explained fundamental theories across multiple systems to help the interdisciplinary research of fake news. The authors presented four main perspectives related to fundamental theories:

- False knowledge it carries

- Different writing styles
- Propagation pattern of Fake News
- The credibility of its creators and spreaders

2.2.2 Investigation of fake news during the 2012 and 2016 U.S. Presidential Election

Allcott et al. (Allcott & Gentzkow 2017) explored the fake news in the 2016 U.S. presidential election's political context. The authors investigated the fake news propagated during the time of elections and their effect on voters. All of these theoretical surveys substantiated the importance of detecting fake news automatically in social media. Fourney et al. (Fourney et al. 2017) explored the trends in fake news consumption during the 2016 U.S. Presidential Election. This relationship observed both at the state level and the county level during the entire election season. It found that fake news circulated during this time-period impacted voters' decision in a quantitative way. Shin et al. (Shin et al. 2018) explored the political misinformation available on social media converging on three main components: temporal pattern, content, and entire information sources. The authors traced a life cycle of 17 widespread political rumours during the 2012 U.S. General Presidential Election. The authors concluded that these rumours affected the image of politicians and polarized the votes.

2.2.3 Investigation of fake news from the prospects of NLP

From NLP perspective, investigation by Young et al. (Young et al. 2018) presented a review of different detection models deployed for numerous NLP tasks. In their research, authors were more focused on the news content-based approaches (Young et al. 2018) for fake news identification. Authors have also explored deep learning models' effectiveness for fake reviews, satire news, and rumours. Oliveira et al. (Oliveira et al. 2021) presented a survey of different methods for pre-processing the data. The future of deep learning in NLP also has a comprehensive survey of recent advances in detecting and mitigating fake news propagating on social media. Features, datasets, and challenges to identify fake news effectively also explored. Anoop et al. (Anoop et al. 2019) did a review of existing fake news detection methods emphasizing content-based features for NLP-based tasksThe authors discussed the techniques to leverage heterogeneous data to control the generation and propagation of fake news that play a fundamental role in fake news detection. The authors discussed datasets, open issues, and future directions. Oshikawa et

TABLE 2.2: Classification results with dataset: Kaggle

Authors	**Accuracy(%)**
(Ghanem et al. 2018)	48.80
(Singh et al. 2017)	87.00
LR-unigram model (Ahmed et al. 2017)	89.00
(Ruchansky et al. 2017)	89.20
(Ahmed et al. 2017) using LSVM model	92.00
(Yang et al. 2018)	92.10
(O'Brien et al. 2018)	93.50

al. (Oshikawa et al. 2020) presented a survey of challenges for automatic fake news detection like meta-data features, and inclusion of temporal features. The authors also discussed datasets, NLP solutions, limits of datasets, and future directions.

2.3 News content-based detection for fake news

Using content-based features, different writing styles (Ghosh & Shah 2018; Mikolov et al. 2013a) can be investigated in addition to feelings or emotions (Liu & Wu 2018; Wang et al. 2018) that found in fake news content. In the existing literature, researchers mainly explored and investigated the detection approaches using content level information (Ghosh & Shah 2018; Shu et al. 2017). Additionally, textual representations were modeled and mainly expressed utilizing tensor factorization method (Shu et al. 2017), deep neural systems (Ghosh & Shah 2018; Liang et al. 2015; Mikolov et al. 2013a; Ruchansky et al. 2017), which performed well to identify fake and real news with real-world fake news datasets. Research using a real-world binary output-label dataset (Kaggle fake news dataset) [†] summarized in table 2.2.

2.3.1 Machine Learning-based models utilizing news content-based features

In this section, existing machine Learning-based approaches presented. In these approaches, authors mainly utilized news content-based features for fake news detection. Different machine

[†] https://www.kaggle.com/c/fake-news/data

learning classifiers investigated and recorded the classification performance with a real-world fake news dataset.

Utilizing linguistic features

In these types of investigations, the researchers mainly employed machine learning-based approaches utilizing linguistic features. Authors primarily investigated the problem of fake news using a binary output-label dataset: Kaggle fake news dataset [‡]. One of the researchers, Perez-Rosas et al. (Pérez-Rosas et al. 2017) explored the automatic classification of fake news on social media. Authors investigated the identification of linguistic features in false news content available on social media. Authors also examined a comparable analysis of the manual identification of fake news. Experiments were conducted utilizing their linguistic-based approach, which achieved an accuracy of 74.00% with a celebrity-news dataset. Ghanem et al. (Ghanem et al. 2018) utilized different word embedding models, including n-gram features, to detect fake articles' stances. The result showed an accuracy of 48.80% using a binary level real-world fake news dataset. In one of the investigations, Singh et al. (Singh et al. 2017) explored the problem of fake news with LIWC features (Linguistic Analysis and Word Count) using machine learning classification models. Authors investigated the problem of fake news with an SVM (support vector machine) model and achieved an accuracy of 87.00% using a real-world fake news dataset. Peters et al. (Peters et al. 2018), one of the researchers carried a different perspective for fake news detection by studying its linguistic features. Despite substantial dependence on linguistic resources, the classifier's performance using a real-world political dataset was even slower than the results done by (O'Brien et al. 2018), with an accuracy of 22.0% only. In one of the study, Monteiro et al. (Monteiro et al. 2018) obtained a fake news dataset in the Portuguese language and examined their results using different linguistic features of news content.

Utilizing TF-IDF and N-gram features

In another research, Castillo et al. (Castillo et al. 2011) examined feature-based approaches to assess the trustworthiness of tweets on Twitter. In their exploration, authors (Ahmed et al. 2017) employed different machine learning algorithms using TF-IDF (Term Frequency and Inverse Document Frequency) as the feature extraction method fake news detection. In another research, (Ahmed et al. 2018) presented a fake news classification method using LR (Linear-regression-based uni-gram model) and achieved an accuracy of 89.00%. Authors have obtained

[‡] `https://www.kaggle.com/c/fake-news/data`

an accuracy of 92% using an LSVM (Linear Support Vector Machine). O'Brien et al. (O'Brien et al. 2018) employed multiple network-based approaches for classifying fake news. Their study achieved an accuracy of 93.50% using the black-box approach for recognizing fake or real news.

Classifying deceptive spam reviews

Ott et al. (Ott et al. 2011) presented a novel approach for identifying deceptive spam reviews on social media. Authors utilized POS tags and word count as features in their research. The authors conducted their experiments using an opinion spam dataset and obtained an accuracy of 90% with their proposed detection method. Feng et al. (Feng et al. 2012) studied the syntactic stylometry for deception detection. The authors also detected deceptive reviews using the rules of Context-Free Grammar (CFG). Authors conducted their experiments based on the review dataset and obtained an accuracy of 91.20%.

Utilizing lexical and syntactic features

Chen et al.(Chen et al. 2015) discussed different lexical, syntactic traits of an article to identify its misleading content. In their investigation, the authors examined a different type of syntactic features for fake news classifications. Bhatt et al. (Bhatt et al. 2018) computed the neural embedding process with a deep recurrent model. The authors finally merged all valuable features and classified the news headline and news body with an accuracy of 43.82%. Rashkin et al. (Rashkin et al. 2017) welcomed a diverse viewpoint on identifying fake news effectively by looking at the news articles' linguistic properties. Authors have investigated fake news with a method: Linguistic Inquiry Word Count (LIWC). The authors have also explored individual words with sentiment lexicon. The authors also analyzed hedging lexicon and lexicons taken from the dataset, despite tangible dependence on linguistic resources, the performance of their model with 22.0% accuracy. Conroy et al. (Conroy et al. 2015) explored with a detailed study about the existing methods for fake news detection. The authors explored the linguistic approach in depth for effective fake news detection with machine learning models. Vedova et al. (Della Vedova et al. 2018) presented a novel method for fake news identification concatenating news content with social context-based features. With their classification approach, an accuracy of 81.70% was achieved utilizing a real-world fake news dataset. Afroz et al. (Afroz et al. 2012) presented different writing style and linguistic features. The authors developed a method to detect stylistic deception in written documents. With their machine learning method, they have achieved an accuracy of 96.6%.

Utilizing news content-based features

Ozbay et al. (Ozbay & Alatas 2020) presented a novel classification approach utilizing an existing real-world fake news dataset. The authors proposed a fake news detection model exploring the news content features as a binary classification problem. Faustini et al. (Faustini & Covões 2020) introduced a novel approach using content text features that can be generated despite source platform and independent of the languages selection. In their research, authors used bag-of-words (BOW) and Word2Vec for the word embedding process. Gravanis et al. (Gravanis et al. 2019) used the dataset (UNBaised-UNB) in their approach , which avoid baize results in the classification task. Study results achieved high accuracy with their designed SVM classifier. Sotirios et al. (Antoniadis et al. 2015) presented a classification model using supervised learning techniques that detect the suspicious pattern of online misinformation. Their model achieved an accuracy of 77% with a real-world fake news dataset.
Rubin et al. (Rubin et al. 2016) presented a conceptual survey of satirical news available on social media. Authors have proposed an SVM-based algorithm that examined their combinations on 360 news articles. With their experimental approach, the authors achieved a precision of 90% and 84% recall. In another research, Vicario et al. (Vicario et al. 2019) presented a classification framework for identifying polarizing content available on social media. The authors used an Italian Facebook dataset and achieved an accuracy of 77% for classifying real and fake news. Shao et al. (Shao et al. 2017) analyzed 14 million messages broadcasting with 400 thousand claims on Twitter during the entire 2016 U.S. presidential campaign and election. Authors have explored the effect of social bots in spreading fake news at a very rapid rate. Chu et al. (Chu et al. 2012) presented an approach to detect human, click-bait and malicious accounts on Twitter. The authors investigated their proposed perspectives utilizing the user-level features to find the likelihood between a human, bot, and malicious user. Their approach demonstrated the effectiveness of their proposed classification method.

2.3.2 Deep Learning-based models for news content-based detection

In this section, existing research work utilizing deep learning approaches presented. In these procedures, authors mainly employed content-based features. Utilizing deep learning-based

networks to identify fake news, Yang et al. (Yang et al. 2018) employed CNN's for better results. The authors investigated the concept of sensitivity analysis in their approach and achieved an accuracy of 92.10%. Liu et al. (Liu & Wu 2018) also examined the methods for identifying false tweets available on social media. Authors utilized a corpus of more than 8 million tweets gathered from the presidential candidates' supporters in the general election in the U.S. where they employed deep CNN's for fake news detection. In their method, authors utilized the theory of subjectivity analysis and obtained an accuracy of 91.00%. Chen et al. (Chen et al. 2018) investigated lexical and syntactic features to identify the news articles' fabricated content. The authors examined potential methods for the automatic detection of click-bait. The authors recommended from investigation that a hybrid approach produce the best results for fake content detection for a binary output label dataset.
In another research, Singhania et al. (Singhania et al. 2017) presented a deep learning-based detector using an attention network. With their approach, authors achieved more accurate results using real-world fake news dataset. Zhao et al. (Zhao et al. 2015) presented a technique based on searching for rumours posts' with inquiry phrases. It combined similar posts, made a cluster, and then deployed the classifier on a cluster group. With their designed system, the authors achieved more accurate results on real-world fake news dataset. Wu et al. (Wu et al. 2017) explored a novel approach for newly emerging rumours using historical data. The authors validated their system using a real-world rumour dataset and demonstrated the performance of their model. It was found that their model performed well on both historical data and new rumours spread on social media. Dong et al. (Dong et al. 2018) presented a hybrid method utilizing news content and side information. The outcome of the investigation showed that attention-based model was more capable for classifying the fake news effectively.

2.4 Social context-based detection of fake news

Social context-based detection approaches are capable of handling the user's features and its connected network in social media. In this section, existing social content-based detection approaches discussed. Capturing the social context present in news articles, the following features can be valuable: user-level features, user post-level features, and network-level features. User-level features were obtained from the user profiles to measure their credibility. (Basak et al. 2019; Potthast et al. 2017). User post-based features highlighted the user's social engagements, propagation path of news stances (Potthast et al. 2017) and the credibility of users (Yang et al. 2019). Network-based features were useful for designing a reliable detection systems for fake news, such as the diffusion networks (Karimi et al. 2018), association networks (Gupta et al. 2018; Yang et al. 2018), and propagation networks (Gupta et al. 2018; Yang et al. 2019; Yang et al. 2018). With the broad adoption of social media networks, research additionally accounted for such online activities for detecting fake news, for example, early detection of fake news by social learning (Mikolov et al. 2013b), user-based relations (Araque et al. 2017), semi-supervised detection (Ren et al. 2016), unsupervised detection (Ren et al. 2016), and also through meta attributes (Giatsoglou et al. 2017; Ren et al. 2016). The authors estimated fake news detection using content and context-level information in their research work, mainly focusing on user comments and check-worthy content (Kim 2014; Kim et al. 2016). In one of the research, authors (Vishwakarma et al. 2019) examined their novel model for veracity analysis of news information and achieved an accuracy of 88.00% with a real-world fake news dataset (FakeNewsNet).

In one of the investigations, Shu et al. (Shu et al. 2019b) studied a method for robotization through hashtag recurrence in the news data. In this study, the authors presented a complete review of psychological and social concepts and existing detection algorithms from a data mining perspective. The results of their research showed further facilitate research on the fake news problem and their directions towards a solution. Zhou et al. (Zhou & Zafarani 2019) investigated four perspectives for analyzing a news article: knowledge-based, style of news writing, propagation of new articles, and credibility of users. Knowledge and style based perspectives employed the news content, whereas propagation-based and credibility-based prospects used the social context. The authors presented a complete overview of the life cycle of fake news and its propagation-based patterns employing social context.

2.4.1 Machine Learning-based detection models utilizing social context

This section illustrated the classification results of numerous implementations conducted using machine learning techniques. Ahmed et al. (Ahmed et al. 2017) examined automatic detection of fake content using available fake reviews on social media. The authors also examined two different feature extraction methods for classifying fake news. The authors experimented with six machine learning models that achieved improved results compared to existing benchmarks. Gupta et al. (Gupta et al. 2012) investigated twitter-based stories' credibility with their novel approach. The authors explored a comparison of their proposed system with the existing methods. Along with that, their research investigated the credibility of news articles with a graph-based optimization approach. Their research also examined a classification approach extracting valuable feature: user-level, tweet-level, and event-based features. Gupta et al. (Gupta et al. 2018) studied a detection method combining content and context-level features in the form of a tensor. With their designed tensor, a tensor factorization method deployed for the classification of the news article. The authors proposed two ideas in their research, News Cohort Analysis and Collaborative News Recommendation for fake news detection. With their results, the authors obtained an F1-score of 81.30%. Zheng et al. (Zheng et al. 2015) investigated a supervised machine learning model for effectively detecting spams on social media. Authors accumulated a dataset from Sina-Weibo [§]. Experiments conducted using an SVM-based (Support Vector Machines) spam detection algorithm. Their result obtained with 99.1% true positive rate for detecting spam reviews.

In one of the research, Natekin et.al. (Natekin & Knoll 2013) employed three neural network architectures for fake news detection with the dataset: "Getting Real about Fake News [¶]" and "Fake News Corpus [‖]" in their investigation. Authors achieved better classification results as compared to existing models using context-related fake news dataset. Zubiaga et al. (Zubiaga et al. 2016) proposed different machine learning models with different word embedding models using the PHEME dataset. With their best approach, authors were able to achieve accuracy with 81%. Sharma et al. (Sharma et al. 2019) investigated the impact of web-based social networking on political decisions. Existing research demonstrated the effect of political discussions and political gatherings. Tacchini et al. (Tacchini et al. 2017) presented a model to identify fake articles based on its user reaction. A bipartite network to classify a news article constructed based on the number of user likes for an article. One of the research, Karimi et al. (Karimi et al. 2018) examined 360 satirical news articles, including civics, science, business, and delicate news. The authors also proposed an SVM-based model. Their proposed framework achieved an accuracy of 38.81%.

2.4.2 Deep Learning-based models for social context-based detection

In this section, deep learning-based approaches discussed from the perspective of social context-based features. In these approaches, authors employed different types of neural networks for effective classification. In recent years, CNN's achieved excellent results in the text representation process for different natural language processing tasks. Spreading fake news has been the point of attention for researchers in recent years, and numerous authors examined this concept

[§] The dataset can be downloaded from: `https://www.researchgate.net/figure/Dataset-from-Sina-Weibo_tbl1_282028558`

[¶] The dataset can be downloaded from: `https://www.kaggle.com/mrisdal/fake-news`

[‖] The dataset can be downloaded from: `https://www.kaggle.com/c/fake-news/data` from various aspects of terminology, sociology, and politics. These approaches were capable of extracting automatic features and improving classification results.

Hybrid models

For deep learning-based approaches, Farajtabar et al. (Farajtabar et al. 2017) presented a multistage intervention framework for handling fake news on social media. By using reinforcement learning, the authors used feature representation of stages, mitigation actions and reward functions. Their method outperformed the existing methods. Ruchansky et al. (Ruchansky et al. 2017) presented a novel hybrid model fusing the source characteristics and user response found in a news data. The authors explored a novel detection approach CSI: Capture, Score, and Integrate. The model experimented on Twitter and Weibo datasets. Their technique was divided into three parts. The first module captured the temporal pattern of user and news article engagement giving its lower dimension representation. The second module indicated a user's credit score based on user features by employing a fully connected layer. The third module combined the vector from the first module and the second module's score to classify the news article effectively. The authors used RNN as a classifier utilizing lower dimension representation of news articles. The authors employed a deep hybrid model for organizing fake news. The authors utilized news-user relationships as an essential factor and achieved an accuracy of 89.20%. Bondielli et al. (Bondielli & Marcelloni 2019) introduced a new hybrid method for classifying the spam messages and reviews by fusing community-based features with other valuable feature as meta-content and user's interaction-based features. Consequently, the authors merged all content and context-based features and classified fake news with an accuracy of 43.82%.

One of the studies, Wang et al. (Wang 2017), introduced a novel dataset in the field of fake news classification. The authors presented a novel architecture to resolve the problem of fake news. The authors built a model using two main components: a Convolutional Neural Network for meta-data learning, followed by a LSTM model subsequently in the designed architecture. Their model, which performed poorly on the test samples, achieved only 27.4% inaccuracy. Roy et al. (Roy et al. 2018) explored the neural embedding approach using a deep recurrent model. The authors used a weighted n-gram bag of word model using statistical features and other external features to feature engineering. The authors explored a combination of CNN and Bi-directional LSTM model. Yang et al. (Yang et al. 2019) explored the possibility of extracting features that lead to fake-news identification by introducing dimensional reduction techniques integrated into the hybrid CNN-LSTM model. Their proposed architecture highlighted the Principal Component Analysis (PCA) and Chi-Square approach, which adopted for feature reduction. In one of the research, Fazil et al. (Fazil & Abulaish 2018) introduced a novel method for identifying the spammers by amalgamating community-based features with other valuable features, namely meta-content and interaction-based features. In one of the research, Ajao et al. (Ajao et al. 2018) identified and classified fake news messages effectively of Twitter posts available on social media with their hybrid neural network model.

Deep learning for rumour detection

In one of the investigations, Allcott et al. (Allcott & Gentzkow 2017) presented a quantitative report to know the impact of fake news on social media in the 2016 U.S. Presidential General Election and its effect upon U.S. voters. The authors examined the authentic and unauthentic URL's related to fake news from the BuzzFeed dataset. Ghosh et al. (Ghosh & Shah 2018) have investigated the impact of web-based social networking on political decisions. In their study, Zhou et al. (Zhou & Zafarani 2018) discussed social media's ability to aggregate a large community of users' judgments. In their further research, the authors explained deep learning approaches with the end aim to develop better rumours detection. Authors examined the difficulties of disseminating rumours, classification of rumours, and related deception methods or frameworks. Authors also reviewed the utilization of such helpful strategies towards designing engaging structures that can help individuals evaluate the integrity of data gathered from various social media platforms. Vosoughi et al. (Vosoughi et al. 2017) identified notable features of rumours by exploring three points of the available information spread online: linguistic writing style and the characteristics of the people involved in propagating information. The authors investigated their proposed algorithm on 209 real-time rumours, including the 2013 Boston Marathon bombings, the 2014 Ferguson unrest, and the 2014 Ebola epidemic. Their study's fundamental objective was to introduce a novel method of assessing style-similarity between different text contents in the form of news. The authors performed the classification using traditional models and obtained an accuracy of 51%. Further, Yang et al. (Yang et al. 2012) appeared with comparative resolutions for detecting false rumours.Throughout the 2011 riots in England, authors noticed that any improvement in the false rumours-based stories could deliver good results. Their investigation of the 2013 Boston Marathon bombings gained some exciting news stories, and most of them were rumours and produced a significant impact on the share market.

In their examination, Zubiaga et al. (Zubiaga et al. 2018a) presented a summary of existing research methods toward social media rumours to develop an effective rumour classification system. Authors have explained the investigation of rumours circulating among nine breaking news occasions. Social media can aggregate the judgments of many users, hence encouraging further study of machine learning approaches to enhance rumours detection frameworks. In a similar type of research, Liang et al. (Liang et al. 2015) compared a novel approach using conditional models with random fields and current rumour detection systems. In this research, the authors explored a novel classifier that improved the precision and recall compared to the existing detection methods. Moreover, the results provided evidence for the generalization of the classifier explained in this research article.

Utilizing recurrent neural network

In another research, Chen et al. (Chen et al. 2018) introduced an unsupervised learning model combining recurrent neural networks and auto-encoders to recognize rumours as anomalies from other credible micro-blogs based on users' behaviours. The experimental results showed that their proposed model obtained an accuracy of 92.49% and an F1-score of 89.16%. Shu et al. (Shu et al. 2017) investigated the relation between fake and real facts available on social media platforms using an open tweet dataset in one similar investigation. This dataset was created by gathering online tweets from Twitter that contains URL from truth checking facts. In their analysis, the authors discovered that URL's are the most widely recognized strategy to share news articles for the measurement of client articulation. Their further research used a Hoax-based dataset that gives a more accurate prediction for distinguishing fake news stories by conflicting them against known news sources from well-known truth inspection sites. In another research, Weiss et al. (Weiss et al. 2020) studied the beginnings of the term "fake news" and the valuable features. This deficiency of consensus has introduced as future suggestions for researchers working in a particular field and higher education. In the deep learning-based investigation, Ma et al. (Ma et al. 2018) applied recurrent neural network (RNN) to obtain contextual features from news articles available online compared to traditional methods that need hand-crafted features. The authors revealed that implementing a deep neural network helped to achieve more accurate results than the existing approaches. Hochreiter et al. (Hochreiter & Schmidhuber 1997) discussed different versions of RNN's to store and access memories. In their investigation, Abedalla et al. (Abedalla et al. 2019) presented a novel framework to detect fake news effectively. The authors developed four different models to validate convolutional layers' performance, LSTM layers, filters, dropout, and batch-normalization process. In their investigation, Agarwal et al. (Agarwal et al. 2016) introduced two datasets for the research community to generate online news reports. The inference is drawn from the overall analysis of their experiments that LSTM outperformed the traditional supervised learning methodologies and delivered a high accuracy.

Utilizing BERT and deep neural network

In one of the investigations, Jwa et al. (Jwa et al. 2019) examined the approach towards automatic fake news detection. The authors employed BERT model to identify fake news by analyzing the relationship between the headline and the news story's body text. Their results improved the F1-score over other existing detection models. Authors also investigated the Twitter-based data of six Venezuelan government officials for tracing fake news. Zhang et al. (Zhang et al. 2018, 2020; Zhang & Wallace 2015) explored a novel architecture (Fake Detector) linking the linguistic and writing-based features obtained from a news article for the classification. Their results demonstrated that their approach was much more practical when compared with existing systems. The authors examined an efficient label propagation algorithm (LPA) for community detection. The authors conducted different experiments on tangible social media interfaces. Their result confirmed that the proposed algorithm was scale-able and exhibits high accuracy. For experimental design, Gated Recurrent Unit (GRU) has used for the extraction of latent features. Further, they have used Gated Diffusive Unit (GDU), combining latent features of news-creators, news articles, and subjects. The authors investigated the problem of fake news using Deep Neural Network with Twitter-based PolitiFact dataset [**] having 14055 tweets with proper fact-check. These news articles refereed to 152 subjects. The authors obtained an accuracy of 63% for binary class interfaces. Collobert et al. (Collobert et al. 2011) deployed neural networks in their research with different convolutional filters to extract global features by max-pooling.

Utilizing convolutional neural network

In their investigation, Crestani et al. (Cerisara et al. 2018) introduced a novel model capable of identifying the different roles of the user as a potential fact checker or a potential spreader. The authors used a CNN architecture and advanced word embedding model to recognize different linguistic patterns in the news data. Shu et al. (Shu et al. 2018) examined a novel neural architecture to detect fake news effectively.. The authors designed a co-attention neural network to

[**] The dataset can be downloaded from `https://twitter.com/PolitiFact`

utilize both news content and user comments. Their recommended method achieved a precision of 90.40% using social-context-based features with the PolitiFact dataset. In another research, Ma et al. (Ma et al. 2016) investigated a neural network utilizing contextual features from news articles. The authors also analyzed fake news with traditional machine learning algorithms, including the performance feedback that requires hand-crafted features. The authors have more concentrated to design effective deep learning methods to achieve better classification results. With their model, they were able to achieve accuracy with 86.12%. In one of the investigation, Yang et al. (Yang et al. 2018) introduced the solution to their novel Convolutional Neural Network (TI-CNN) with a combination of both text and image-based features using fusion techniques. Their approach did a detailed investigation using the news article's visual content for effective classification with some based content-based features. Explicit features from textual content include linguistic, psychological perspective, lexical diversity and sentiment analysis.
In a similar type of research, Shu et al. (Shu et al. 2019b) investigated the problem of fake news with their proposed model (TriFN). Their approach was capable of handling the interactions between the user and news articles. It also delivers the best understanding between credible and non-credible news publishers. Experiments conducted with two real-world fake news datasets (BuzzFeed and PolitiFact) and achieved an accuracy of 86.40% and 87.80%, respectively. Zhou et al. (Zhou & Zafarani 2019) investigated different news pattern present in social media news. Their approach studied the patterns available in online social networks responsible for the fake news propagation rapidly. The authors achieved an accuracy of 83.50%. In one of the investigation, Wang et al. (Wang et al. 2018) investigated a novel framework known as Event Adversarial Neural Network (EANN) utilizing event-invariant features. Experiments were conducted with two large scale real-world datasets: Twitter and Weibo. The authors also investigated many deep learning methods using a Kaggle fake news dataset and authenticated news articles from Signal Media News. The authors recognized that LSTM, GRU (Gated Recurrent Unit), and Bi-LSTM-based (Bi-directional LSTM) classifiers produced better outcomes than CNN-based models. Experiments carried out using two real-world fake news datasets (Weibo and Twitter) and achieved an accuracy of 71.50% and 82.70%, respectively.
In their investigation, Zhou et al. (Zhou et al. 2020) investigated a unique approach-SAFE, which studied multi-modal information (both textual and visual) available on social media. Experiments on two real-world datasets (PolitiFact and Gossip-cop) achieved an accuracy of 87.40% and 83.80%, respectively. Zhong et al. (Zhong et al. 2019) have examined a fast Gaussian kernel learning method by solving a specifically structured global optimization problem in

similar research. They have used the improved Hoffman's approximation method in their examination. The authors produced good classification performance with their proposed approach. Kim et al. (Kim 2014) have proposed a CNN model with multiple filters and the different size of the input filter. Kalchbrenner et al. (Kalchbrenner et al. 2014) proposed a novel method for text classification. Bhatt et al. (Bhatt et al. 2018) proposed an automated process for identifying news articles' veracity through deep learning techniques in their investigation. The overall accuracy delivered by their CNN architecture with the unevenly distribute FNC-1 dataset was 74.84%.

Utilizing graph convolutional neural network

In one of the research, Wang et al. (Wang et al. 2021) explored a graph convolutional method concatenating with an LSTM layer before the output integrates local sequential order and global semantic relationship-relationship recognition of fake news available in the form of text. The proposed model, SemSeq4FD, has shown a minimum improvement of 1.17% over existing and current models, with the highest accuracy, resulted in 93.7% on the LUN dataset. In another research, Dharawa et al. (Dharawat et al. 2020) released Covid-HeRA, a new misinformation dataset to identify the fake news related to COVID-19. Their proposed CNN model used an advanced embedding model- Glove for input word length vectors. The testing accuracy achieved using a binary classification model categorizing the news articles was 96.6%. Orso et al. (Orso et al. 2020) have emphasized that the diffusion of social media leads to increasing the degree of clarity in sharing scientific information, which has led to the spreading of inaccurate data from reliable sources at the beginning of the COVID-19. Gupta et al. (Gupta et al. 2012) evaluated the credible-score of Twitter events with a Page-rank-like credibility approach. Comparison of classifier approach, fundamental credit analysis and event graph-based optimization approach has presented. The event graph-based optimization approach provides better results as it employs event similarity.

Our research contributions

To detect fake news better, multiple studies were conducted and reported their results as research contributions. In the preceding research work (Kaliyar et al. 2020b), the problem of fake news with their designed deep convolutional neural network employing content-based features investigated. It was found that the proposed convolutional model obtained better classification outcome as compared to existing benchmarks. Further, a deep neural network with four dense layers was designed (Kaliyar et al. 2020a) using both contents and context-based features. It achieved improvable results using the dataset: BuzzFeed and PolitiFact. In this research, it was recommended that the deep neural network is one step ahead, with a better classification performance on a binary output dataset, as compared to existing systems. In another study, Kaliyar et al. (Kaliyar et al. 2021b), a novel BERT-based in-depth learning approach was explored for fake news classification. A multichannel-CNN architecture was explored by using an advanced word embedding technique. In the architecture, three parallel channel of 1D-CNN's were combined into a unified structure. This combination was effective because multiple-branch convolution networks different filters and kernel sizes for effective feature learning with variable-sized input documents. It observed that the proposed model is a general framework that can be used with any representation models in bi- nary fake news categories. The model considered different filter sizes across each dense layer with varying length feature mapping capability. The accuracy of the designed system was highest among all the existing detection models. A preliminary version of this work appeared in the CODS-COMAD-2021 (Kaliyar et al. 2021b).

2.5 Research gaps in literature

After going through various existing literature that created a body of knowledge in fake news detection techniques, the following research gaps were identified based on the comprehensive literature investigation (refer tables 2.3-2.6) :

2.5.1 Deficiency of context-dependent datasets

Critical estimation of the literature revealed that existing datasets contain news content, social context of news, and user-related learning. These information were valuable for fake news propagation, detection, and mitigation in an effective manner. To the best of our knowledge (Ahmed et al. 2017; Basak et al. 2019; Fazil & Abulaish 2018; Kumar & Shah 2018; Ruchansky et al. 2017) existing datasets only contained one or two aspects.

2.5.2 Improvisation in existing fake news detection techniques

As mentioned in the existing literature review, the accuracy of existing fake news detection methods lied in the range of 44-80% using real-time context-specific datasets. An enormous scope of improvement had been promising (Ghosh & Shah 2018; Potthast et al. 2017; Shu et al. 2017; Wason 2018; Zhou & Zafarani 2018). The end goal of this research work was to utilize

TABLE 2.3: Summary of Existing work using News Content-based features

Dataset	**Classification Approach**	**Features**	**Classifier and Accuracy (%)**
Twitter Dataset (Fazil & Abulaish 2018)	Machine learning-based (Hybrid model)	Content, and interaction-based features	RF, DT, NB (94.70)
FNC-1 dataset (Ahmed et al. 2017)	Deep learning-based (Neural Network)	Neural, statistical, and external features	CNN, LSTM (43.82)
BuzzFeedWebis and FakeNewsCorpus 2016 (Potthast et al. 2017)	Machine learning-based	Salient features of rumors	RF (75.00)
Fake.Br Corpus (Monteiro et al. 2018)	Machine learning-based	Content level features	SVM (89.00)
LIAR (Wang 2017)	Machine learning-based (Multi-class Fake News Detection framework)	Content features and meta features	SVM, RF, MMFD (38.81)
LIAR (Wang 2017)	Deep learning-based (Hybrid convolutional neural network)	Content features	LR, SVM (27.40)
PolitiFact (Rashkin et al. 2017)	Deep learning-based	300	NB and LSTM (22.00)
2013 Boston Marathon bombings and 2014 Ebola epidemic (Vosoughi et al. 2017)	Machine learning-based	Content and salient features of rumors	LR, SVM, HMM(75.00)
Celebrity-news (Pérez-Rosas et al. 2017)	Machine learning-based	Linguistic Features	74.00
(Ghanem et al. 2018)	Machine learning-based	News content-based features (n-gram features)	48.80
(O'Brien et al. 2018)	Deep learning-based	Content-based features	NN(Black-box) 93.50
Spam-reviews (Ott et al. 2011)	Machine learning-based	Content-based features	90.00
(Feng et al. 2012)	Machine learning-based	Linguistic and Content-based features	91.20
(Rashkin et al. 2017)	Machine learning-based	Linguistic and Content-based features	22.00

TABLE 2.4: Summary of Existing work using News Content-based features

Dataset	**Approach**	**Features**	**Classifier and Accuracy (%)**
(Antoniadis et al. 2015)	Machine learning-based	Content, and interaction-based features	77.00
(Rubin et al. 2016)	Machine learning-based	News content-based features	SVM (90.00-precision and 84.00 recall)
(Della Vedova et al. 2018)	Machine learning-based	News content-based features	81.70
Italian Facebook dataset (Vicario et al. 2019)	Machine learning-based	Content level features	77.00
(Yang et al. 2018)	Deep learning-based	News content and User-based features	96.60
(Liu & Wu 2018)	Deep learning-based	Content and linguistic-based features	NN, CNN (91.00)

TABLE 2.5: Summary of existing work using Social-context-based features

Dataset	**Approach**	**Features**	**Classifier and Accuracy (%)**
PolitiFact (Gupta et al. 2018)	Machine learning-based	News content and Social context-based features	SVM (81.30)
Sina-webio (Zheng et al. 2015)	Machine learning-based	Spam detection using News Social-based features	TPR
PHEME (Zubiaga et al. 2018a)	Machine learning-based	Social context-based features	81.00
(Ruchansky et al. 2017)	Neural network	Social context-based features	HNN (89.20)
(Vosoughi et al. 2017)	Deep learning approach	Salient features of rumours (linguistic style, characteristics of users, and network patterns)	51.00
(Chen et al. 2015)	Deep learning-based	Micro-blogs based on users' behaviors	RNN (89.16 F1-Score)
LIAR (Roy et al. 2018)	Deep learning-based approach	Social context-based features	CNN and BI-LSTM (43.82)
(Shu et al. 2019d)	Deep learning-based approach	Social-context-based features	CNN, LSTM(90.40)

TABLE 2.6: Summary of existing work using Social-context-based features

Dataset	**Approach**	**Features**	**Classifier and Accuracy (%)**
Twitter-based PolitiFact (dataset) (Zhang et al. 2018)	Deep learning-based	News content and Social context-based features	DNN (63.00)
(BuzzFeed and PolitiFact) (Shu et al. 2019b)	Deep learning-based	News content and Social context-based features	86.40 and 87.80
(BuzzFeed and PolitiFact) (Zhou & Zafarani 2018)	Deep learning-based	News content and Social context-based features	83.50
(Weibo and Twitter) (Wang et al. 2018)	Deep learning-based	Social context-based features	EANN (71.50 and 82.70)
(PolitiFact and Gossipcop) (Zhou et al. 2020)	Deep learning-based	News content and Social context-based features	87.40 and 83.80
(Ma et al. 2016)	Deep learning-based	News content and Social context-based features	NN (86.12)
(Bhatt et al. 2018)	Deep learning-based approach	Social context-based features	CNN and BI-LSTM (74.84)

news content and social-context features effectively. A significant challenge in predicting the fake news article was the diverse writing style to deceive society's people. News content was not only sufficient for better classification results on various real-world fake news datasets. Apart from traditional machine learning methods, an effective detection model using deep learning techniques was necessary. The proposed designed models gave satisfactorily result in this regard with many real-world fake news datasets.

2.5.3 To build a model or framework that combines multiple features (content, context, community, and temporal, etc.)

There existed three principal segments of the fake article: the text of an article, the acknowledgment of user experiences, and the source of fake creators developing it. Existing research work primarily enlighten toward one or two components, which has limited their success and generality. Scope to establish a model or framework that combines all these characteristics and other prospects for a more accurate and automated prediction was promising—most of the existing works revolved around news content-based classification methods. However, from an industrial and academic point of view, context and community-based detection were more significant. Nevertheless, there were present a few studies on the community-based detection approaches.

2.5.4 Fake News Detection utilizing different user-cohorts

It was also identified that the user does not exist in an isolation on online social media (Kumar & Shah 2018). User's generally form a cohort/community among themselves to share their views. Community structure comes from the theory of equally like-minded users tend to follow each other. Scope to use these exiting communities (Gupta et al. 2018) on social media for fake news detection was promising.

2.5.5 To build a in-depth learning approach utilizing a bi-directional word embedding model

Based on the existing studies, it was observed that there was a primary necessity for an effective deep learning approach utilizing the power of bi-directional pre-trained word embedding like BERT for more accurate results. Most of the existing methods (Fazil & Abulaish 2018; Kumar & Shah 2018; Zhou & Zafarani 2018) mainly used a single direction pre-trained model GloVe and Word2Vec.

2.5.6 Generalization

For the detection of fake news, existing classification techniques (Ghosh & Shah 2018; Shu et al. 2017; Zhou & Zafarani 2018) were limited to specific social-network data and a language. A generalized model to classify fake articles was the primary necessity to achieve accurate results with any real-world news dataset. The fundamental motivation behind developing the proposed model was to produce effective learning on distinct fake news datasets. The capability of the designed model was lucrative towards any real-world fake news dataset with reasonable good accuracy.

2.6 Objectives of the present work

With a comprehensive literature investigation and acknowledging the research gaps, the following research objectives were chosen for further research:

- RO-1: To Investigate and analyze the existing fake news detection techniques.
- RO-2: Acquisition and creation of context-dependent real-world datasets for fake news detection.
- RO-3: To customize the existing fake news detection methods using deep learning techniques for improvisation.
- RO-4: Verification and Validation of the customized techniques on the new real-world datasets.
- RO-5: Comparative analysis of the proposed framework in terms of classification accuracy, F1-score, and confusion matrix.
- RO-6: Generalized results based on the experiments performed on various real-world fake news datasets.

2.7 Research methodology

For achieving the above-mentioned research objectives, the following research methodologies were adopted:

- An exploration and investigation of existing fake news detection techniques (RO-1): A detailed investigation of the existing detection methods done successfully. All existing detection methods were investigated for conceding the current processes in the research. In conclusion, a broad scope of improvement was present.

- Acquisition and creation of context-dependent datasets (RO-2): We acquired the real-world fake news datasets like Real-or-fake (Shape of the dataset: 6335*4), and attributes are: id, title, text, label, FakeNewsNet (Attributes are: id, URL, title, text, tweet_ids etc.), Kaggle-Fake-news (Training data: 20800 and Testing data: 5800) and attributes are: id, title, author, text, and label, PHEME (70000 instances), LIAR (six output classes), CoAID-19 (In this dataset, there was a total of 2138 instances of news with 549 of fake in total, and it was the first publicly available dataset for COVID-19 related to fake news articles). One novel fake news dataset was built named: FN-COV (A novel dataset in which there are 69,976 instances of news with 44.84% of fake in total). The collection had several topics like COVID-19, quarantine, and social distancing tag related news articles.

- Installation and study of the tool(s)(RO-1 and RO-3): Python toolkit with NumPy, Pandas, and scikit-learn/Keras version 2 with either a Theano or Tensor-Flow at the back-end for the implementation of the existing/proposed methods for fake news identification.

- To customize the existing fake news detection methods using deep learning techniques for improvisation (RO-3, RO-4, and RO-5): An effective deep learning model (FNDNet) was designed. With the proposed model (FNDNet), more accurate results were achieved than current detection methods. Subsequently, a deep learning model (DeepFakE) was designed utilizing news content, and social context features in a 3-mode tensor. A CMTF method used to convert a higher-dimensional vector into lower-dimensional vectors. To validate the proposed model's performance, two real-world fake news datasets (BuzzFeed and PolitiFact) used in the research. More improved results were achieved with the proposed approach. Subsequently, moving towards the following research objective, a BERT-based in-depth learning approach (FakeBERT) was developed with a bi-directional pre-trained word embedding model (BERT). More accurate results were obtained as compared to existing benchmarks using a real-world fake news dataset. Afterwards, a deep learning model was designed for the effective identification of fake news. In this approach, news content, social context, and user-level features were combined with user-cohorts in the form of the 3-mode tensor. More accurate results were achieved with the proposed model with both the dataset: BuzzFeed and PolitiFact. The proposed model's performance was validated and verified employing different performance parameters.

- Generalization (RO-6): Moving towards generalized results, a hybrid multichannel convolutional neural network was designed to detect fake news effectively. The performance of our proposed model was validated with three real-world fake news datasets. A multichannel convolutional neural network was designed with different kernel-size convolutional layers and filters for better learning. The authors used an embedding layer in traditional models, accompanied by a 1D-CNN having one pooling layer and a prediction output layer. The proposed model combined multiple 1D-CNN's that read the source document with different kernel sizes. As a result, the input document processed at different n-grams at a moment. The proposed model also learned how to combine these studies (different sized n-grams) best and how it affects model learning.

2.8 Organization of the thesis

Fundamental theories behind fake news detection presented in the current chapter. The present chapter also summarized the work available in the literature and the scope of improvement. Recent research has undertaken to understand the proposed deep neural networks' performance with other existing benchmarks. The following chapter included a description of the deep neural networks for effective fake news detection. Chapter 3 represented the results corresponding to our designed BERT-based deep learning approach and FNDNet (the proposed deep neural network). Chapter 4 described the results corresponding to the Tensor decomposition-based in-depth learning approach and EchoFakeD (the proposed deep neural network handling multiple types of user-based features). Chapter 5 represented the results corresponding to the hybrid model for effective fake news detection for binary classification. Chapter 6 described the generalized results corresponding to the designed deep learning approach with three novel and real-world fake news datasets. Chapter 7 represented the conclusion and future scope.

Chapter 3

Improving Fake News Detection with FNDNet and BERT-based deep learning approach*

*The results presented in this chapter are published in: 1. **Kaliyar, R. K.**, Goswami, A., Narang, P., & Sinha, S. (2020). FNDNet–a deep convolutional neural network for fake news detection. Cognitive Systems Research, 61, 32-44. https://doi.org/10.1016/j.cogsys.2019.12.005.**(Elsevier, SCI Impact Factor: 1.902)** .
2. **Kaliyar, R. K.**, Goswami, A., & Narang, P. (2021) FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. Multimedia Tools and Applications, 1-24.https://doi.org/10.1007/s11042-020-10183-2.**(Springer Nature, SCI Impact Factor: 2.313).**

In the current chapter, fake news identification with FNDNet and the BERT-based deep learning approach presented.

3.1 FNDNet: A Deep convolutional neural network

This section represented the research work using the proposed model (FNDNet) for effective fake news detection.

3.1.1 Introduction

This chapter represents the research work with the proposed model (FNDNet) for fake news detection with a more in-depth convolutional approach. A look around exhibited that our recommended system did not rely on obtaining hand-crafted features. Alternatively, the model (FNDNet) outlined learning discriminatory features through deep learning automatically. The proposed model's architecture was encouraged by recent advancement in the field of fake news identification (Fu et al. 2017; Pan et al. 2018; Ruchansky et al. 2017; Zhang & Wallace 2015). The model showed excellent performance on large-scale real-world fake news datasets compared to existing fake content detection methods. CNN's achieved excellent performance in many text classification tasks and different industrial applications (Liang et al. 2015; Roy et al. 2018; Zhou & Zafarani 2018). The proposed model would also help provide a better solution for such industrial applications in business, retails and insurance, etc. The proposed model also addressed choosing an optimal depth of CNN's for the text classification problem effectively. The classification results outperformed the current models for fake news detection. The model decreased classification error. In this research, the proposed model's performance was demonstrated for language representation, where accuracy of 98.36% has achieved. Using the proposed method, improved results were obtained compared to the baseline approaches, made FNDNet a promising model for accurate fake news detection.

3.1.2 Experimental setup and methodology

In this section, various experiments and techniques presented that were utilized to achieve the research objectives and the results of the experiments.

Word embedding

The primary advantage of using word embedding models (Caliskan et al. 2017; Camacho-Collados et al. 2016; Peters et al. 2018; Wang et al. 2016; Zhang & Wallace 2015) was the ability of training with massive real-world datasets. Embedding generally symbolized geometrical encoding (Zhang et al. 2015) of words based on how frequently they appear together in a text corpus. It reduced the time consumption for training the model, cleaning, and processing. This research displaced the processing layer parameters with input embedding vectors for using pre-trained embedding models for training. Primarily, the index was kept and then fixed this layer, restricting it from being updated throughout gradient descent (Peters et al. 2018; Qi et al. 2018). The experiments have shown programs that embedding-based input vectors perform a valuable role in text classification tasks. Pre-prepared models can be categorized into two forms, viz., context-free and contextual-based. Contextual-based models can further be divided as unidirectional or bidirectional (Cerisara et al. 2018; Iyyer et al. 2015; Kamkarhaghighi & Makrehchi 2017) for pre-training.

GloVe

The GloVe is an unsupervised learning algorithm (Ahmed et al. 2017; Asparouhov & Muthén 2010; Cerisara et al. 2018; Hailong et al. 2014; Medhat et al. 2014) that mainly uses to discover the closeness of two words in the form of a vector space. These generated vector representations are called word embedding vectors. In GloVe, training performs on aggregated global word-word co-occurrence matrices (Ahmed et al. 2017) to from a corpus. This research work used the most delicate word embedding package is 822Mb, called "glove.6B.zip". The 100-dimensional version was used in our research. GloVe gave lower weight for widespread word pairs to prevent meaningless stop words such as "the", "an", etc., which did not dominate the training progress. Before training the model, a co-occurrence matrix X created based on input words, where a cell X_{ij} was a "strength", which defines how often the word i appears in the context of the word j. Once X is ready, it is important to choose our input vector for each word in the dataset.

$$w_i^T \tilde{w}_j + b_i + \tilde{b}_j = logX_{ij} \tag{3.1}$$

where b_i and b_j are scalar bias terms associated with words i and j, respectively.

The end goal was to minimize the target objective function J, which was helpful in recording all the squared errors, weighted with a function f:

TABLE 3.1: FNDNet layer Architecture Model

Layer (type)	Input size	Output size
Embedding	1000	1000 × 100
Conv1d	1000 × 100	998 × 128
Conv1d	1000 × 100	997 × 128
Conv1d	1000 × 100	996 × 128
Maxpool	998 × 128	199 × 128
Maxpool	997 × 128	199 × 128
Maxpool	996 × 128	199 × 128
Concatenate	199 × 128, 199 × 128, 199 × 128	597 × 128
Conv1d	597 × 128	593 × 128
Maxpool	593 × 128	118 × 128
Conv1d	118 × 128	114 × 128
Maxpool	114 × 128	3 × 128
Flatten	3 × 128	384
Dense	384	128
Dense	128	2

$$J = \sum_{i,j=1}^{V} f(X_{ij})(w_i^T \tilde{w}_j + b_i + \tilde{b_j} - logX_{ij})^2 \tag{3.2}$$

Where V is defined as the size of the vocabulary.

Importance of pre-trained Word Embedding

In the recent time, word embedding techniques have been broadly employed in classification tasks using textual data. The high accuracy with pre-trained word embedding models significantly influenced fake news classification for large datasets. For pre-trained embedding experiments, the processing layer's parameters were displayed with pre-trained embedding vectors, maintaining the index and preventing it from being updated during the process of gradient descent. The primary findings were illustrated in terms of (i) training loss, (ii) confusion matrix, and (iii) accuracy. The experiment demonstrated that word embedding-based vectors played an influential role in fake news detection.

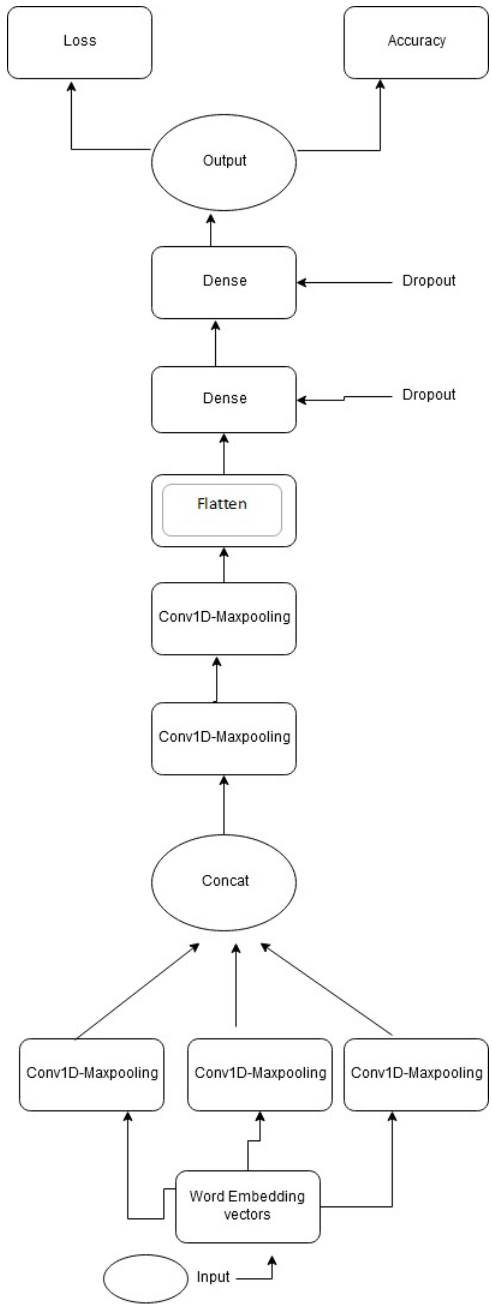

FIGURE 3.1: Computational Graph for FNDNet Model

The proposed deep convolutional network: FNDNet

Figure 3.1 showed the computational flow of the proposed model FNDNet. In most of the existing researches (Cerisara et al. 2018; Zhong et al. 2019), fake news detection has been investigated by adopting a standard text classification model that consists of an embedding layer as input in the form of word embedding vectors (Zhong et al. 2019), followed by a one-dimensional CNN (Zhong et al. 2019). The proposed model's design was motivated by the concept of multiple parallel channels-variable-size-based neural networks (Zhong et al. 2019). The proposed model reaped the benefits of traditional feature engineering and automated feature engineering (Seide et al. 2011). In the proposed model, inputs were the word-embedding vectors generated from GloVe. The same input word embedding vectors were provided to all three parallel convolutional layers. Subsequently, the FNDNet elements, viz., the choices of number of convolutional layers, different kernels sizes in each convolutional layer, number of dense layers, dropout, and the selection of activation function were discussed to make the proposed detection model more efficient and deep convolutional-based, as follows:

- **Convolutional layer:** It is defined as the core functional block of any CNN-based networks for classification (Zhong et al. 2019). This layer consists of a set of filters or kernels (Zhong et al. 2019). With these filters' help, a tiny part of the input data is taken at a time for processing and applied across the whole input. The fundamental operations performed by this convolutional layer are matrix multiplications-based operations (Zhong et al. 2019), which passes through an activation function (Zhong et al. 2019) to produce the final output. In the proposed model, three parallel convolutional layers were used with different kernel sizes. The primary motivation behind it was to contain more information in different word vectors during training.

- **Max-pooling layer:** A pooling layer effectively down-samples (Nagi et al. 2011; Zhong et al. 2019) the last layer's output in a neural network, reducing the number of operations required for all the following layers present in the network. In the proposed architecture, three max-pooling layers consolidated the output from the convolutional layers. Two more convolutional layers were taken, followed by max-pooling layers, to make the proposed architecture deeper convolutional-based to achieve efficient results.

- **Flatten layer:** Flatten layer is denoted by a function that transforms the input features and maps them to a single column for further processing. One flatten layer was used in the outlined model.

- **Dense layer or fully connected layer:** A dense layer's functionality is considered a linear operation (Zhong et al. 2019) in which every input is connected to every output by some weight. Two dense layers were taken to make the proposed model dense. Researchers (Vasudevan et al. 2019; Zhong et al. 2019) used one or two dense layers before the final softmax layer in their research work.

- **Dropout:** Dropout is a regularization technique (Vasudevan et al. 2019; Zhong et al. 2019), which aims to reduce the complexity of any classification model with the end goal to prevent over-fitting (Zhong et al. 2019). Dropout to all fully-connected layers/dense layers was employed. Dropout at each layer of the network showed promising results. The value of dropout was taken as 0.2 throughout our experiments.

- **Activation Function:** In this research, ReLU (Rectified Linear Unit) as a activation function (Li & Yuan 2017) was used. ReLU also enhances the nonlinear properties (Li & Yuan 2017) of the complete network's decision-making function without modifying the receptive fields (Li & Yuan 2017) of the convolution layer. It is the most commonly used activation function in Deep Learning due to efficient results. It is computationally efficient than sigmoid or Tanh and solves the vanishing gradient problem. The equation of ReLU can be written as:

$$\sigma = max(0, z) \tag{3.3}$$

In Table 3.1, the layered structure of the designed FNDNet model tabulated. In the architecture, the input was distributed into three parallel convolutional layers having 128 kernels. The first convolution layer has 128 kernels of size 3, which decreases the input vector from 1000 to 998; the second convolution layer has 128 kernels of size 4, which decreases input vector from 1000 to 997; and the third convolution layer has 128 kernels of size five which decreases input vector from 1000 to 996 after convolution. After each convolution layer, a max-pooling layer is present to reduce the dimension. Following this, the max-pooling layer has filter size 5, which further reduces the vector to 1/5th of 996, i.e. 199. After concatenation, a convolution layer applied of size 5 with 128 kernels followed by a max-pooling layer. Finally, it followed by a dense hidden layer of 128 neurons. The output of the FNDNet model receives through a dense layer with a dropout value of 0.2.

Importance of Deeper CNN

In the modern era of computing, neural networks with 100 or more layers exist. Neural networks have trained using back-propagation and forward-propagation algorithms. In these algorithms, the gradient (derivative) of the cost function is used to update each layer's weights. The gradient's value decreases with each new layer, primarily when the sigmoid activation function is used. This problem is also known as the vanishing gradient. Direct connection in dense or deeper CNN solves this problem. Deeper CNN is also less prone to over-fitting as compared to standard CNN.

3.1.3 Experimental Results & Analysis

Dataset description

Experiments conducted using the fake news dataset [†]. It consists of two files (i) train.csv: A full training dataset (refer Tables 3.2 for more details), and (ii) test.csv: A dataset without the output label (refer Table 3.2 for more details). This dataset is associated with the fake articles spread during the time of the 2016 U.S. Presidential Election. In this dataset, ID represents the unique value for a particular news article; the title represents the main heading of news; the author expresses the news creator's name. Text is the main core part of this dataset, representing the complete news article, and labels provide information about the article as potentially unreliable or reliable.

Hyperparameter Tuning

The process of choosing hyperparameters is a necessary aspect of any deep learning classification model. Hyperparameters are the variables that set before applying a learning algorithm to a context-specific dataset. For selecting and optimizing hyperparameters, there were two basic approaches: manual and automatic selection. Both methods were technically viable. The decision typically outlined a trade-off between the deep understanding of the model required to select hyperparameters manually versus the high computational cost required by automatic selection algorithms. Tables 3.3-3.5 listed the hyperparameters' values in the designed experiments for achieving accurate classification.

[†]. The dataset can be downloaded from: `https://www.kaggle.com`

TABLE 3.2: Fake News dataset

Attribute	Number of Instances
ID	20800
title	20242
author	18843
text	20761
label	20800

TABLE 3.3: Hyperparameters for CNN

Hyperparameter	Value
No. of convolution layers	3
No. of max pooling layers	3
No. of dense layers	2
Loss function	Categorical-crossentropy
Activation function	ReLU
Learning rate	0.001
Optimizer	Ada-delta
Number of epochs	5
Batch size	128

TABLE 3.4: Hyperparameters for LSTM

Hyperparameter	Value
No. of convolution layers	2
No. of max pooling layers	2
No. of dense layers	4
Dropout rate	.2
Optimizer	Adam
Activation function	ReLU
Loss function	Binary cross-entropy
Number of epochs	10
Batch size	64

TABLE 3.5: Hyperparameter for FNDNet.

Hyperparameter	Value
No. of convolution layers	5
No. of max pooling layers	5
No. of dense layers	4
Dropout rate	.2
Optimizer	Adadelta
Activation function	ReLU
Loss function	Categorical cross-entropy
Number of epochs	5
Batch-size	128

3.1.4 Performance parameters

Precision, recall, F_1-Score, TNR, FPR, and accuracy were used as evaluation matrices to assess the proposed model's performance. To control the different embedding types, the following hyper-parameters were fixed (refer to Tables 3.3-3.5) throughout the experiments.

Confusion Matrix

A confusion matrix represents the information about actual and predicted classifications performed by a classifier.

Precision & Recall

The measure of the model's ability to accurately identify the occurrence of a positive class instance was determined by recall as:

$$Recall = \frac{TP}{TP+FN} \tag{3.4}$$

where precision is :

$$Precision = \frac{TP}{TP+FP} \tag{3.5}$$

TABLE 3.6: Confusion Matrix for MNB

	Predicted Positive	**Predicted Negative**
Actual Positive	898 (TP)	73 (FN)
Actual Negative	111 (FP)	853 (TN)

F_1-Score

F_1 score can be calculated as:

$$F_1 = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \tag{3.6}$$

Specificity or True Negative Rate (TNR)

TNR can be calculated as:

$$TrueNegativeRate(TNR) = \frac{TN}{FP + TN} \tag{3.7}$$

False Positive Rate (FPR)

FPR is calculated as:

$$FalsePositiveRate(FPR) = \frac{FP}{FP + TN} \tag{3.8}$$

Accuracy

Accuracy can be defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{3.9}$$

where True positive (TP) = correctly identified
False positive (FP) = incorrectly identified
True negative (TN) = correctly rejected
False negative (FN) = incorrectly rejected

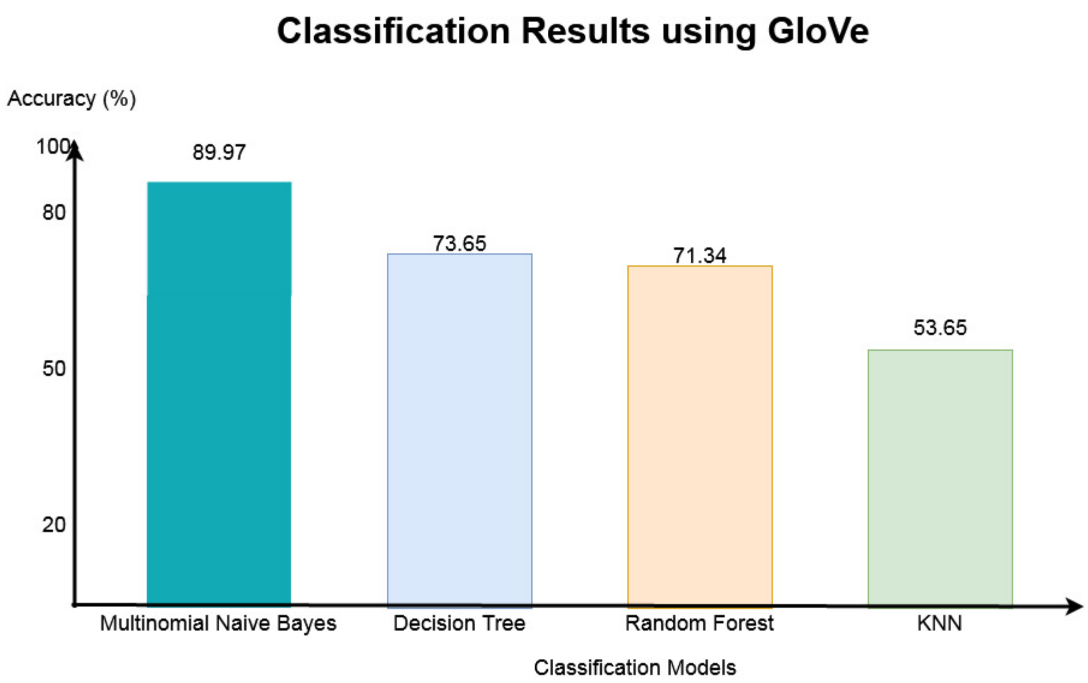

FIGURE 3.2: Machine learning-based Classification results using GloVe

TABLE 3.7: Confusion Matrix for KNN

	Predicted Positive	**Predicted Negative**
Actual Positive	836 (TP)	200 (FN)
Actual Negative	762 (FP)	282 (TN)

TABLE 3.8: Confusion Matrix for DT

	Predicted Positive	**Predicted Negative**
Actual Positive	901 (TP)	135 (FN)
Actual Negative	413 (FP)	631 (TN)

TABLE 3.9: Confusion Matrix for RF

	Predicted Positive	**Predicted Negative**
Actual Positive	802 (TP)	234 (FN)
Actual Negative	361 (FP)	683 (TN)

TABLE 3.10: Confusion Matrix for CNN with GloVe

	Predicted Positive	**Predicted Negative**
Actual Positive	882 (TP)	76 (FN)
Actual Negative	90 (FP)	952 (TN)

TABLE 3.11: Confusion Matrix for LSTM with GloVe

	Predicted Positive	**Predicted Negative**
Actual Positive	995 (TP)	47 (FN)
Actual Negative	8 (FP)	1030 (TN)

TABLE 3.12: Confusion Matrix for FNDNet with GloVe

	Predicted Positive	**Predicted Negative**
Actual Positive	995 (TP)	32 (FN)
Actual Negative	6 (FP)	1018 (TN)

TABLE 3.13: Classification results using Machine Learning and Deep Learning-based models

Embedding Model	**Classification Model**	**Precision(%)**	**Recall(%)**	**F_1-Score(%)**
Tf-Idf on unigrams and bigrams	Neural Network	95.31	92.78	94.03
BoW without unigram and bigrams	Neural Network	91.45	87.67	89.57
Word2Vec	Neural Network	80.23	72.34	76.08
GloVe	Mutinomial Naive Bayes	88.99	92.48	90.70
GloVe	Decision Tree	68.56	86.97	73.68
GloVe	Random Forest	68.95	77.41	72.94
GloVe	KNN	52.31	80.69	63.37
GloVe	CNN	90.74	92.07	91.40
GloVe	LSTM	99.20	95.49	97.31
GloVe	**The Proposed model (FNDNet)**	**99.40**	**96.88**	**98.12**

TABLE 3.14: True Negative Rate (TNR) and False positive rate (FPR)

Word Embedding Model	Classification Model	TNR(%)	FPR(%)
Tf-Idf on unigrams and bi-grams	Neural Network	85.12	14.52
BoW without unigram and bi-grams	Neural Network	62.17	12.23
Word2Vec	Neural Network	59.35	37.65
GloVe	Mutinomial Naive Bayes	88.49	11.51
GloVe	Decision Tree	60.44	39.56
GloVe	Random Forest	65.42	34.58
GloVe	KNN	27.01	72.98
GloVe	CNN	91.36	8.64
GloVe	LSTM	99.22	0.77
GloVe	**The Proposed model (FND-Net)**	**99.41**	**0.59**

TABLE 3.15: Classification results using Machine Learning and Deep Learning-based models

Word Embedding Model	Classification Model	Accuracy (%)
Tf-Idf on unigrams and bi-grams	Neural Network	94.31
BoW without unigram and bi-grams	Neural Network	89.23
Word2Vec	Neural Network	75.67
GloVe	Mutinomial Naive Bayes	89.97
GloVe	Decision Tree	73.65
GloVe	Random Forest	71.34
GloVe	KNN	53.75
GloVe	CNN	91.50
GloVe	LSTM	97.25
GloVe	**The Proposed model (FND-Net)**	**98.36**

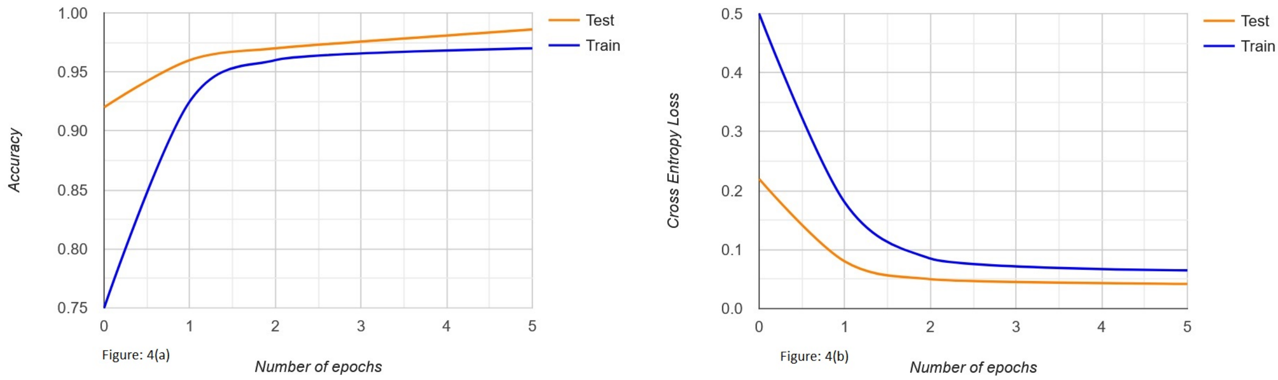

FIGURE 3.3: Accuracy & Model loss or Cross-entropy loss of FNDNet model with Training and Testing samples

Results

Firstly, numerous experiments were conducted to evaluate different machine learning classifiers' performance using a real-world fake news dataset (refer to Table 3.2 for more details). Respective confusion matrices (Table 3.6-3.12) chosen for each machine learning as well as the deep learning-based classifier for evaluating the performance using different estimating parameters (for more details, refer to section 5.2). This research found that using Multinational Naive Bayes (MNB) as a classifier achieved an accuracy of 89.97%. Machine Learning-based classification results are shown in Figure 3.2. MNB consists of a higher actual negative rate and less false positive rate among all machine learning-based models. Performance decreases as the scale of data increases during the investigation with machine learning-based models. Accuracy was also low in machine learning-based classifiers due to limited handwritten features present in the respective dataset. Facing these problems, deep-learning-based implementations were motivated. Deep learning automatically finds out the essential features for classification, whereas in Machine Learning, the features that were fed manually. Deep neural networks with different pre-trained word embedding models were implemented, such as Word2Vec, BOW, etc. We achieved better accuracy (94.31%) as compared to previous machine learning-based models with GloVe. The end goal of this research work was to develop a more effective model for fake news classification.

Considering the issues in machine-learning based implementations, deep learning-based models (CNN, LSTM), and the proposed model (FNDNet)) were implemented and recorded their performances for fake news identification. It found that using the GloVe-enabled pre-trained word embedding technique with CNN, a training accuracy of 64.50% was obtained, with a

validation accuracy of 91.50%, with five epochs. LSTM using GloVe was implemented and recorded a training accuracy of 99.74% and a testing accuracy of 97.25% with ten epochs. The proposed model was executed (FNDNet) using the GloVe pre-trained word embedding model and recorded a testing accuracy of 98.36%. It showed that the best accuracy was obtained by the designed deep learning model. Tables 3.13-3.14, compiled the values of different performance parameters (refer to Table 3.13) and (true negative rate and false-positive rate in Table 3.14) for machine learning as well as deep learning-based classification models. From Figure 3.3, the accuracy and cross-entropy loss of the implemented deep CNN-based model with training and testing data values can be seen. It observed that with the increasing number of epochs, testing accuracy increased, and model loss reduced significantly with the proposed model.
From the above investigation, as seen in Figure 3.3, the training loss decayed quickly with the GloVe-enabled deep convolutional neural network compared to the standard embedding model. Table 3.1, the training loss utilizing advanced pre-trained word embedding models decayed comparatively fast and without any fluctuations. Figure 3.3 demonstrated that cross-entropy loss reduced significantly using the FNDNet model. The recommended model obtained the highest accuracy and minimal losses.

3.1.5 Comparative analysis with existing results

To explore the proposed model's effectiveness (refer to Table 3.16 for more details), a comparative analysis outlined with the existing methods utilizing the Kaggle fake news dataset. The highest benchmark was reported with an accuracy of 93.50% for fake news detection using the corresponding dataset. Also, from Table 3.15, the proposed model showed comparatively better results and effectiveness (training accuracy, testing accuracy, lightweight model for training). Using the GloVe-enabled deep convolutional-based approach, an accuracy of 98.36% was achieved. The proposed model achieved better results (refer to Table 3.15 for more details) with real-world text-based fake news datasets compared to existing works.

3.2 BERT-based deep learning approach

In this section, research work with the proposed BERT-based Deep learning approach presented.

TABLE 3.16: A comparison of implemented results with the dataset: Kaggle

Authors	Accuracy(%)
(Ghanem et al. 2018)	48.80
(Singh et al. 2017)	87.00
Using LR-unigram model (Ahmed et al. 2017)	89.00
(Ruchansky et al. 2017)	89.20
(Ahmed et al. 2017) using LSVM model	92.00
(Yang et al. 2018)	92.10
(O'Brien et al. 2018)	93.50
FNDNet	**98.36**

3.2.1 Introduction

In the existing approaches (De Sarkar et al. 2018; Ghosh & Shah 2018; Pérez-Rosas et al. 2017), many valuable methods presented using traditional learning standards. The fundamental advantage of using a deep learning model over existing classical feature-based approaches was that it did not require any handwritten features; instead, it identifies the best feature set on its own. Deep CNN's powerful learning ability was primarily needed due to complicated feature extraction that can automatically learn different language representations. In the existing approaches (Fazil & Abulaish 2018; Roy et al. 2018; Tenney et al. 2019), several inspiring ideas presented to bring advancements in deep Convolutional Neural Networks (CNN's) like employing temporal and graph channel level information, depth of architecture, and graph-based multi-path information processing. The purpose of using a block of layers as a structural unit was also obtaining notoriety among researchers. The proposed method recommended a BERT-enabled system (FakeBERT) utilizing the power of the advanced word embedding-BERT. BERT was employed as a sentence encoder, which ideally got the context representation of a sentence. This work contrasts with previous research works (De Sarkar et al. 2018) where researchers looked at a text sequence in a unidirectional way (either left to right or right to left for pre-training). Many existing and practical methods had been (De Sarkar et al. 2018; Malik et al. 1991) presented with sequential neural networks to encode the relevant information. However, a deep neural network with a bidirectional training approach can be an optimal and accurate solution for detecting fake news. The proposed method improved fake news classification performance capturing semantic and long-distance dependencies in sentences.

A classification layer was added on the top of the encoder output to design the proposed architecture, multiplying the output vector by the embedding matrix and finally calculating each vector's probability with the Softmax function. The model consolidated three parallel blocks of 1D-convolutional neural networks, with BERT having different kernel sizes and filters followed by a max-pooling layer across each block. With this combination, the documents processed utilizing different CNN-based systems with varying kernel size (different n-grams), filters, and several hidden layers. The design of FakeBERT consists of five convolution layers, five max-pooling layers followed by two densely connected layers and one embedding layer (BERT-layer) of input. In each layer, several filters have applied to extract the information from the training dataset. Such a combination of BERT with a one-dimensional deep convolutional neural network (1d-CNN) helps handle large-scale structure and unstructured text. It effectively addresses ambiguity, which was the most significant challenge to natural language understanding. Experiments conducted to validate the outcome of the proposed model. Numerous performance evaluation parameters (training accuracy, validation accuracy etc.) have been taken into consideration to validate the classification results. Extensive experimentation demonstrated that the proposed model outperformed compared to the existing benchmarks for classifying fake news. The performance of our bidirectional pre-trained model (BERT) was illustrated and achieved an accuracy of 98.90%. The proposed approach produced improved results by 4% comparing to the baseline approaches and was promising for detecting fake news on real-world datasets.

3.2.2 Experimental setup and methodology

In current section, an overview of the existing word embedding (refer to figure 3.4 for more details), GloVe word embedding, BERT model, fine-tuning BERT processes, and the selection of optimal hyperparameters discussed. In this section, numerous experiments and procedures presented that were utilized to achieve the research objectives along with the results of the experiments.

BERT

BERT (Devlin et al. 2018) is a advanced pre-trained word embedding technique based on transformer encoded architecture (Tenney et al. 2019). BERT as a sentence encoder was employed, which can accurately obtain the context representation of a sentence. BERT removed the unidirectional constraint using a mask language model (MLM) (Tenney et al. 2019). It randomly masked some of the tokens from the input and predicted the original vocabulary id of the masked

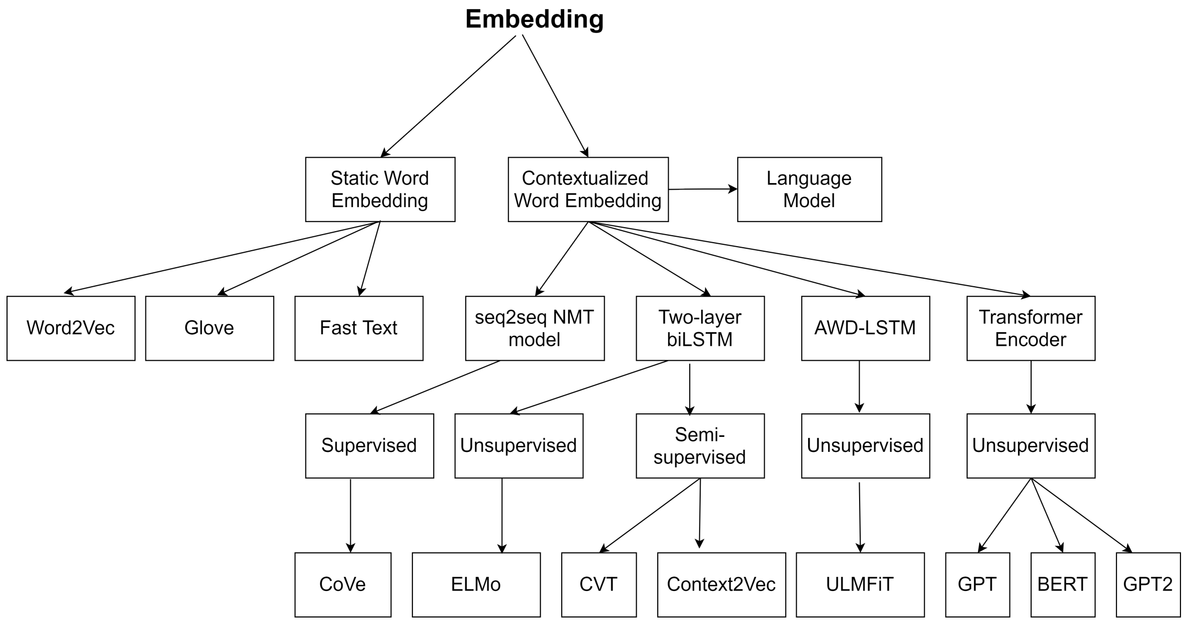

FIGURE 3.4: An Overview of existing word-embedding models

word based only. MLM has increased the capability of BERT to outperforms as compared to previous embedding methods. In this research, an embedding vector for a sentence was extracted or a set of words or pooling hidden states' sequence for the whole input sequence. A deep bidirectional model was more powerful than a shallow word embedding model. In the existing research (Devlin et al. 2018), two types of BERT models investigated for context-specific tasks were:

- BERT Base (refer table 3.17 for more information about parameters setting): Smaller in size, computationally affordable and not applicable to complex text mining operations.
- BERT Large (refer table 3.18 for more information about parameters setting): Larger, computationally expensive, and crunches extensive text data to deliver the best results.

Fine-tuning of BERT

Fine-tuning of BERT (Devlin et al. 2018) is a process that enables the model of many downstream tasks, irrespective of the text pattern (single text or text pairs). A limited exploration was available to enhance BERT's computing power to enhance the classification performance of target tasks. BERT model uses a self-attention mechanism to unify the word vectors as inputs that include bidirectional cross attention between two sentences. Mainly, a few fine-tuning strategies need to consider: 1) The first factor is the pre-processing of long text since the maximum

TABLE 3.17: Parameters for BERT-Base

Parameter Name	Value of Parameter
Number of Layers	12
Hidden Size	768
Attention Heads	12
Number of Parameters	110M

TABLE 3.18: Parameters for BERT-Large

Parameter Name	Value of Parameter
Number of Layers	24
Hidden Size	1024
Attention Heads	16
Number of Parameters	340M

sequence length of BERT is 512. In this research, the sequence length of 512 was taken. 2) The second factor is layer selection. The proposed BERT-based approach was designed to consist of an embedding layer, a 12-layer encoder, and a pooling layer. 3) The third factor was the over-fitting problem. BERT can be fine-tuned with different learning parameters for different context-specific tasks (Tenney et al. 2019) (refer table 3.18 for more information).

Deep learning methods for fake news detection

Deep learning models are well-known for achieving better results in a broad spectrum of artificial intelligence applications (Qi et al. 2018). Current section presented a summary of the proposed research utilizing deep learning models with their architectures to achieve the performance goal. Experiments conducted using different deep learning-based models (CNN and LSTM (Greff et al. 2016)) and the proposed model (FakeBERT) with another pre-trained word embedding (BERT and GloVe):

- Convolutional Neural Network (CNN): In figure 3.5, the proposed designed CNN model's computational graph showed. This CNN model truncated, zero-pads, and tokenizes the fake news article independently and passes each into an embedding layer. In this architecture (refer table 3.19 and figure 3.5), first convolution layer holds 128 filters with

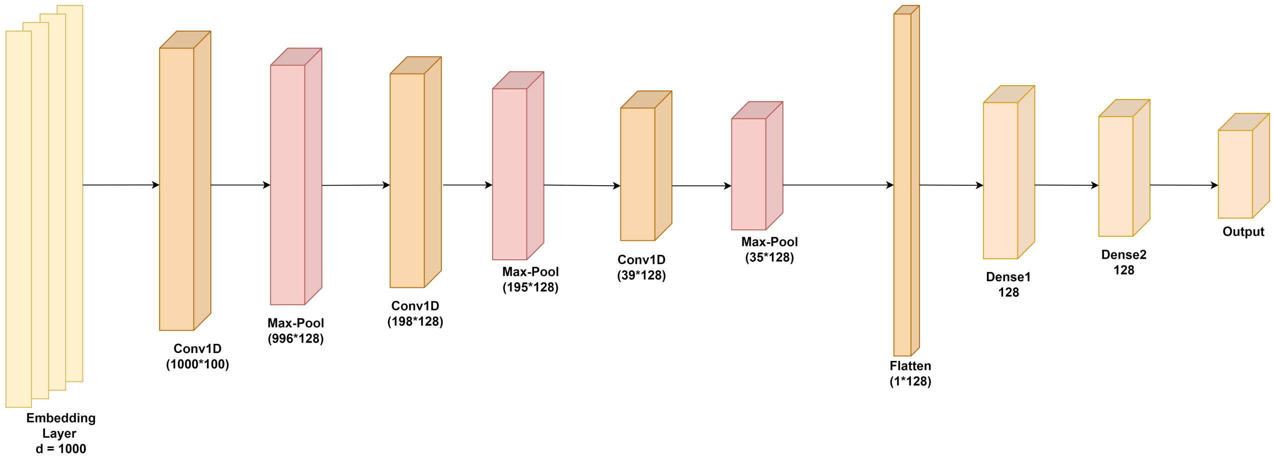

FIGURE 3.5: CNN Model

kernels_size=5, which decreases the input embedding vector from 1000 to 996 after convolution process. A max-pooling layer was also present after each convolution layer to reduce the network's input vector dimension. Subsequently, a max-pooling layer with filter_size=5; that further minimizes the embedding vector to 1/5th of 996, i.e. 199. The second convolution layer holds 128 filters with kernels_size=5, which decreases the input embedding vector from 199 to 195. Subsequently, this was the max-pooling layer with filter size 5; that further reduced the input vector to 1/5th of 199, i.e. 39. After three convolution layers, a flatten layer added to convert 2-D input to 1-D. Subsequently, two hidden layers having 128 neurons in each one. The CNN's outputs were passed through a dense layer with dropout and then passed through a softmax layer to yield a stance classification. Several trainable parameters showed in table 3.19.

- LSTM model: In this investigation, the LSTM model was implemented with four dense layers with a batch normalization process to classify fake news. From table 3.20, the layered architecture was presented of the LSTM model.

The proposed deep learning approach: FakeBERT

The most fundamental advantage of selecting a deep learning model was an automatic feature extraction capability. With the proposed model, the input was passed in the form of a tensor. More concrete results were achieved with a deep architecture that consists of hierarchical representations of learning. From figure 3.6, the proposed approach's computational graph (FakeBERT) showed. In many existing and helpful studies (Cerisara et al. 2018; Zhong

TABLE 3.19: CNN layered architecture

Layer	Input size	Output size	Param number
Embedding	1000	1000 × 100	25187700
Conv1D	1000 × 100	996 × 128	64128
Maxpool	996 × 128	199 × 128	0
Conv1D	199 × 128	195 × 128	82048
Maxpool	195 × 128	39 × 128	0
Conv1D	39 × 128	35 × 128	82048
Maxpool	35 × 128	1 × 128	0
Flatten	1 × 128	128	0
Dense	128	128	16512
Dense	128	2	258

TABLE 3.20: LSTM layered architecture

Layer	Input size	Output size	Param number
Embedding	1000 × 100	1000 × 100	25187700
Dropout	1000 × 100	1000 × 100	0
Conv1D	1000 × 100	1000 × 32	16032
Maxpool	1000 × 32	500 × 32	0
Conv1D	500 × 32	500 × 64	6208
Maxpool	500 × 64	250 × 64	0
LSTM	250 × 64	100	66000
Batch-Normalization	100	100	400
Dense	100	256	25856
Dense	256	128	32896
Dense	128	64	8256
Dense	64	2	130

et al. 2019), the problem of fake news examined utilizing a unidirectional pre-trained word embedding model followed by a 1D-convolutional-pooling layer neural network (Seide et al. 2011; Zhong et al. 2019). In the proposed model, inputs were the vectors generated after the word-embedding process from BERT. The equal dimensional were given input vectors to all three convolutional layers present in parallel blocks (Munandar et al. 2018) followed by a pooling layer in each block. The decision of choosing several convolutional layers, kernels_sizes, no. of filters, and optimal hyperparameters etc.(Guo et al. 2019; Munandar et al. 2018) to make the proposed system more accurate was as follows:

- **Convolutional layer:**. The proposed model used three parallel blocks of 1D-CNN with one layer in each block and two straight forward layers after the concatenation process with different kernel sizes and filters.

- **Max-pooling layer:** The proposed model used five max-pooling layers (three using parallel blocks of 1D-CNN and two with straight forward convolutional layers).

- **Flatten layer:** In between the convolutional layer and the fully connected layer, there was a Flatten layer. Flattening transforms a two-dimensional matrix of features into a vector fed into a fully connected neural network classifier.

- **Dense layer:** A dense layer is considered as a regular layer of neurons in a neural network. In many existing and practical methods (Seide et al. 2011; Vasudevan et al. 2019), authors primarily used one or two dense layers in their proposed networks to prevent over-fitting. In the proposed model, two dense layers were taken with a diverse number of filters.

- **Dropout:** The dropout to dense layers was applied in the proposed network. Dropout worked by randomly setting the outgoing edges of hidden units to 0 at each update of the training phase. The value of dropout was used as 0.2 in the investigations.

- **Activation Function:**. The equation of ReLU (Li & Yuan 2017; Sibi et al. 2013) can be written as:

  $$\sigma = max(0, z) \tag{3.10}$$

  here z =input

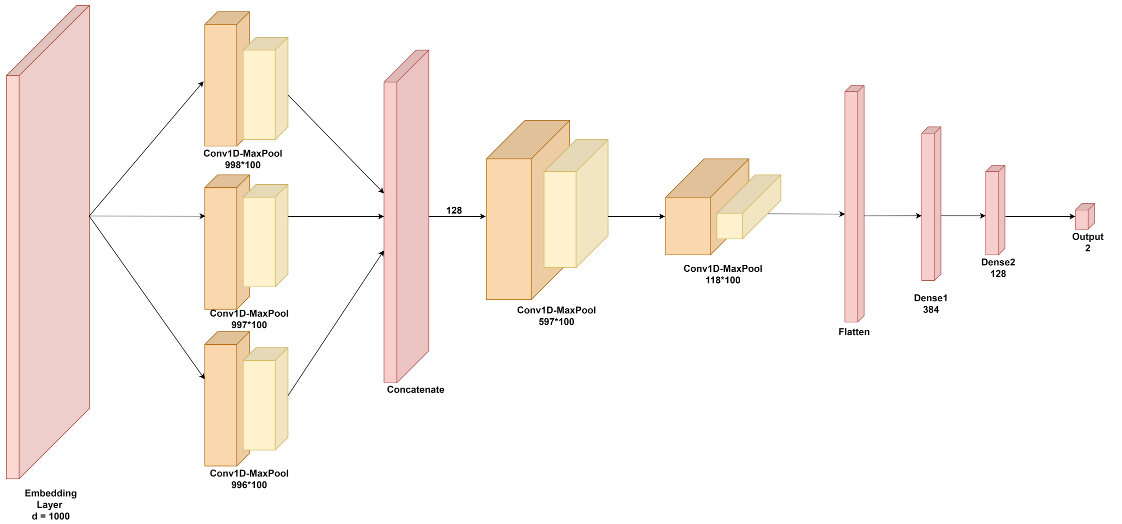

FIGURE 3.6: FakeBERT Model

- **Loss Function (L):** The cross-entropy goes down as the prediction gets more and more accurate. It becomes zero if the prediction is perfect. In binary classification, cross-entropy can be calculated as:

$$L = -(ylog(p) + (1-y)log(1-p)) \tag{3.11}$$

  Here y - binary indicator (0 or 1), p - predicted probability

The computational graph and layered architecture of the proposed FakeBERT model showed using table 3.21 and figure 6.5. In this design, the input was divided into three parallel blocks of 1D-CNN, having 128 filters and one convolutional layer across each block. The first convolution layer consists of 128 filters and kernel_size=3, which decreased the input embedding vector from 1000 to 998; the second layer has 128 filters and kernel_size=4, which reduced the input vector from 1000 to 997. The third layer has 128 filters and kernel_size=5, which decreased the input vector from 1000 to 996. After a particular convolution layer, a max-pooling layer was also present to decrease the dimension. Subsequently, a max-pooling layer with kernel_size=5 further reduces the vector to 1/5th of 996, i.e. 199. After concatenating the three above Conv-layers, a convolution layer was applied to have kernel_size=5 including 128 filters. Subsequently, two hidden layers were having 384 and 128 nodes, respectively. The number of trainable parameters across each layer was also presented (for more details, refer to column Param number) in table 3.21. The proposed model was not computationally complex for training at any real-world fake news dataset. The work was carried using the NVIDIA DGX-1 V100

TABLE 3.21: FakeBERT layered architecture

Layer	Input size	Output size	Param number
Embedding	1000	1000 × 100	25187700
Conv1D	1000 × 100	998 × 128	38528
Conv1D	1000 × 100	997 × 128	51328
Conv1D	1000 × 100	996 × 128	64128
Maxpool	998 × 128	199 × 128	0
Maxpool	997 × 128	199 × 128	0
Maxpool	996 × 128	199 × 128	0
Concatenate	199 × 128, 199 × 128, 199 × 128	597 × 128	0
Conv1D	597 × 128	593 × 128	82048
Maxpool	593 × 128	118 × 128	0
Conv1D	118 × 128	114 × 128	82048
Maxpool	114 × 128	3 × 128	0
Flatten	3 × 128	384	0
Dense	384	128	49280
Dense	128	2	258

machine. The machine was equipped with 40600 CUDA cores, 5120 tensor cores, 128 GB RAM and 1000 TFLOPS speed.

3.2.3 Experiments

Experiments conducted using deep learning models (CNN and LSTM) and the proposed model (FakeBERT) using pre-trained word embedding techniques (BERT and GloVe). Performances were recorded of different classification models and analyzed with the benchmark results.

Dataset description

In this research, extensive experiments were conducted using the real-world fake news dataset [‡]. It (refer table 3.25) consists of two files (i) train.csv, and (ii) test.csv: A testing dataset without the label. It was a collection of fake and real news propagated during the U.S. General Presidential Election-2016. Table 3.26 showed the class labels' instances in the respective fake news dataset.

[‡] The dataset can be downloaded from: https://www.kaggle.com

TABLE 3.22: Optimal Hyperparameters with CNN

Hyperparameter	**Value**
Number of convolution layers	3
Number of max pooling layers	3
Number of dense layers	2
Number of Flatten layers	1
Loss function	Categorical cross-entropy
Activation function	ReLU
Learning rate	0.001
Optimizer	Ada-delta
Number of epochs	10
Batch size	128

TABLE 3.23: Optimal Hyperparameters with LSTM.

Hyperparameter	**Value**
Number of convolution layers	2
Number of max pooling layers	2
Number of dense layers	4
Dropout rate	.2
Optimizer	Adam
Activation function	ReLU
Loss function	Binary cross-entropy
Number of epochs	10
Batch size	64

TABLE 3.24: Optimal Hyperparameters with FakeBERT.

Hyperparameter	**Value**
Number of convolution layers	5
Number of max pooling layers	5
Number of dense layers	2
Number of Flatten layers	1
Dropout rate	.2
Optimizer	Adadelta
Activation function	ReLU
Loss function	Categorical cross-entropy
Number of epochs	10
Batch size	128

TABLE 3.25: Attributes in the Fake News dataset

Attribute	**Number of Instances**
ID	20800
title	20242
author	18843
text	20761
label (information about that the news as fake or real)	20800

Hyperparameter setting

Existing deep learning models explicitly defined optimal hyperparameters that examine several factors such as memory and cost. The optimal selection of best numbers depends on the balanced or imbalanced dataset. From Tables 3.22-3.24, the values of hyperparameters used in the proposed approach observed.

Evaluation Parameters

To evaluate the performance of FakeBERT, accuracy, cross-entropy loss, FPR (False Positive Rate), and FNR (False Negative Rate) were considered evaluation matrices.

TABLE 3.26: Fake News dataset with the class labels

Class label	Number of Instances
True	10540
False	10260

TABLE 3.27: Confusion Matrix for MNB with GloVe

	Predicted negative	Predicted positive
Actual negative	853 (TN)	111 (FP)
Actual positive	73 (FN)	898 (TP)

3.2.4 Results and discussion

The classification outcomes were examined and analyzed with several classifiers utilizing different learning paradigms (optimal hyper-parameters and architectures). Classification results demonstrated that automatic feature extraction with deep learning models played an essential role in detecting fake news more accurately. The proposed model (FakeBERT) generated more accurate results than existing benchmarks with an accuracy of 98.90%.

Classification results using machine learning models

Firstly, numerous experiments were conducted for evaluating the performance of different machine learning classifiers (Multinomial Naive Bayes (MNB), Random Forest (RF), Decision Tree (DT), K-nearest neighbor (KNN)) using a real-world fake news dataset. The examination found that using MNB; we achieved an accuracy of 89.97% with GloVe. Confusion matrices with machine learning classifiers showed in tables 3.28- 3.30. The decision-tree algorithm also provided an accuracy of 73.65%. The confusion matrix using the MNB classifier predicts more

TABLE 3.28: Confusion Matrix for KNN with GloVe

	Predicted negative	Predicted positive
Actual negative	282 (TN)	762 (FP)
Actual positive	200 (FN)	836 (TP)

TABLE 3.29: Confusion Matrix for DT with GloVe

	Predicted negative	**Predicted positive**
Actual negative	631 (TN)	413 (FP)
Actual positive	135 (FN)	901 (TP)

TABLE 3.30: Confusion Matrix for RF with GloVe

	Predicted negative	**Predicted positive**
Actual negative	683 (TN)	361 (FP)
Actual positive	234 (FN)	802 (TP)

TABLE 3.31: Confusion Matrix for LSTM with GloVe

	Predicted negative	**Predicted positive**
Actual negative	1030 (TN)	8 (FP)
Actual positive	47 (FN)	995 (TP)

TABLE 3.32: Confusion Matrix for CNN with BERT

	Predicted negative	**Predicted positive**
Actual negative	1004 (TN)	63 (FP)
Actual positive	90 (FN)	942 (TP)

TABLE 3.33: Confusion Matrix for LSTM with BERT

	Predicted negative	**Predicted positive**
Actual negative	1032 (TN)	7 (FP)
Actual positive	44 (FN)	998 (TP)

TABLE 3.34: Confusion Matrix for FakeBERT with BERT

	Predicted negative	Predicted positive
Actual negative	1045 (TN)	6 (FP)
Actual positive	17 (FN)	1012 (TP)

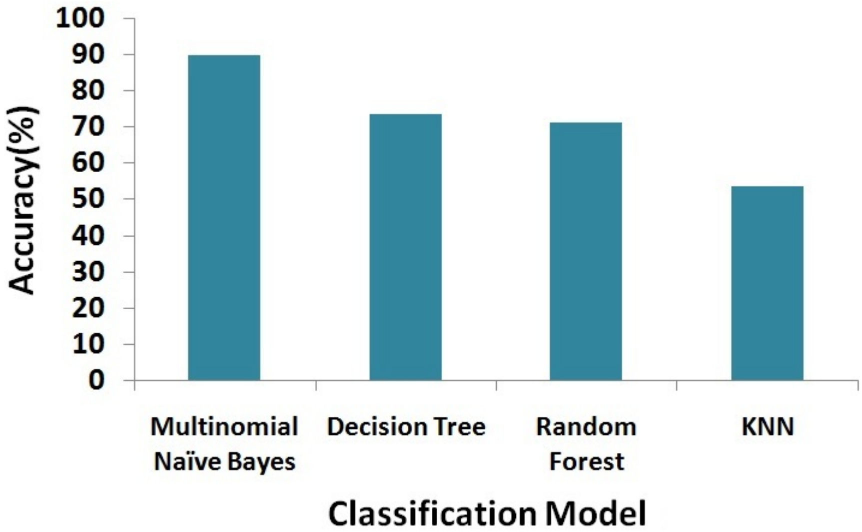

FIGURE 3.7: Classification results with GloVe

labels accurately closer to actual labels with the testing dataset (for more details, refer to table 3.27). Machine Learning-based classification results tabulated in table 3.36 and figure 3.37. The investigation found that accuracy was not up to the mark with real-world fake news dataset. Further, a bidirectional training model, a more robust feature extractor (Tenney et al. 2019) was on priority for study. Motivated by this fact, BERT-a bidirectional transformer encoder-based pre-trained word embedding model was introduced. BERT was a more robust feature extractor than GloVe and provided effective results for NLP tasks. Experiments conducted using the BERT-based machine learning approach and achieved improved classification results.

Classification results using deep learning models

To improve the classification results and considering the issues in machine learning approaches, more further experiments conducted with the deep learning-based models (CNN, LSTM, and FakeBERT) and recorded the performances with many real-world fake news datasets. A deep convolutional network was designed with BERT as a word embedding model. The proposed

FIGURE 3.8: Accuracy and Cross Entropy Loss using CNN

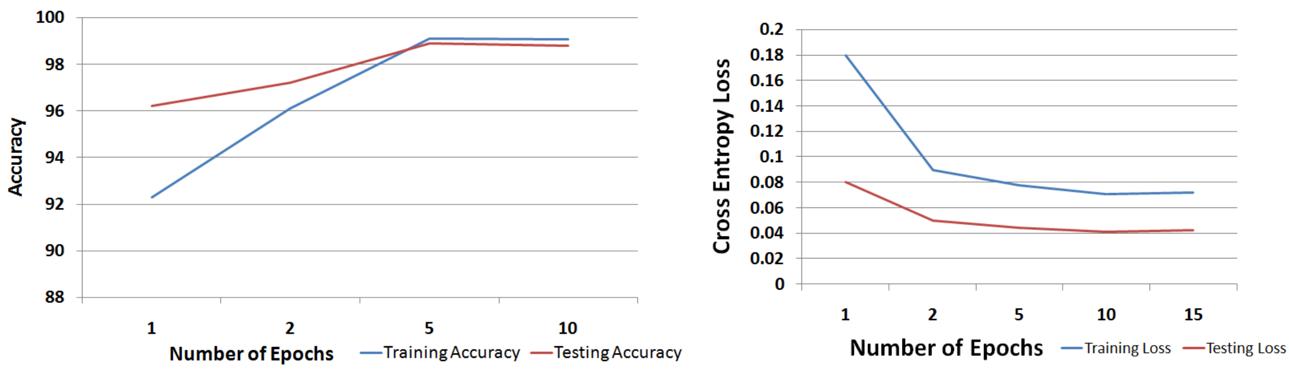

FIGURE 3.9: Accuracy and Cross Entropy Loss using FakeBERT

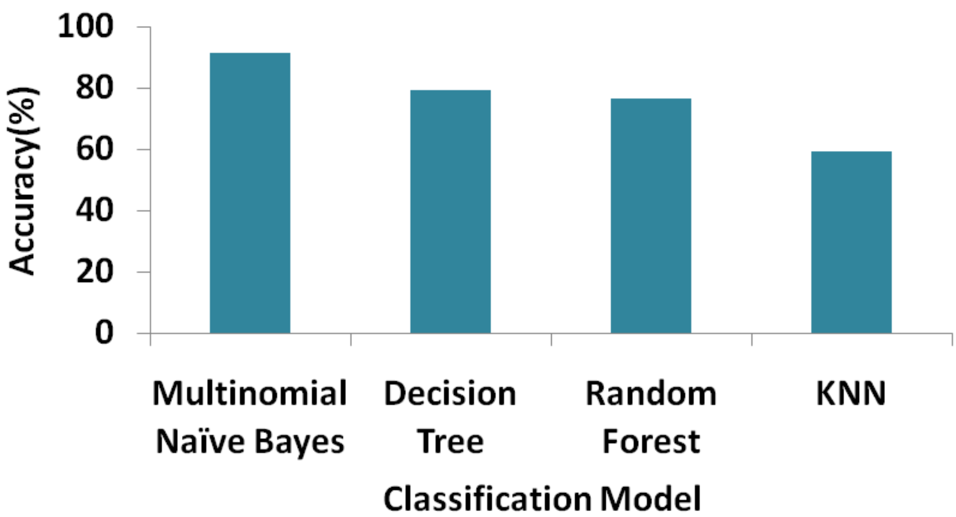

FIGURE 3.10: Classification results with BERT

TABLE 3.35: The proposed model vs existing benchmarks with real-world fake news dataset

Authors	Accuracy(%)
(Ghanem et al. 2018)	48.80
(Singh et al. 2017)	87.00
(Ahmed et al. 2017) using LR-unigram model	89.00
(Ruchansky et al. 2017)	89.20
(Ahmed et al. 2017) using LSVM model	92.00
(Liu & Wu 2018)	92.10
(O'Brien et al. 2018)	93.50
The Proposed model (FakeBERT)	**98.90**

TABLE 3.36: Classification results with BERT and GloVe

Word Embedding Model	Classification Model	Accuracy (%)
TF-IDF (using unigrams and bigrams)	Neural Network	94.31
BOW (Bag of words)	Neural Network	89.23
Word2Vec	Neural Network	75.67
GloVe	MNB	89.97
GloVe	DT	73.65
GloVe	RF	71.34
GloVe	KNN	53.75
BERT	MNB	91.20
BERT	DT	79.25
BERT	RF	76.40
BERT	KNN	59.10
GloVe	CNN	91.50
GloVe	LSTM	97.25
BERT	CNN	92.70
BERT	LSTM	97.55
BERT	**Our Proposed model (FakeBERT)**	**98.90**

TABLE 3.37: Values of FPR and FNR

Word Embedding Model	**Classification Model**	**FPR**	**FNR**
TF-IDF (using unigrams and bigrams)	Neural Network	0.04684	0.0742
BOW (Bag of words)	Neural Network	0.1040	0.0862
Word2Vec	Neural Network	0.1320	0.3416
GloVe	MNB	0.1151	0.0752
GloVe	DT	0.3956	0.1303
GloVe	RF	0.3458	0.2259
GloVe	KNN	0.7299	0.1931
BERT	MNB	0.0985	0.0789
BERT	DT	0.1660	0.2429
BERT	RF	0.1245	0.3318
BERT	KNN	0.4037	0.4110
GloVe	CNN	0.0989	0.0776
GloVe	LSTM	0.0080	0.0482
BERT	CNN	0.0590	0.0872
BERT	LSTM	0.0077	0.0451
BERT	**FakeBERT**	**0.0160**	**0.0059**

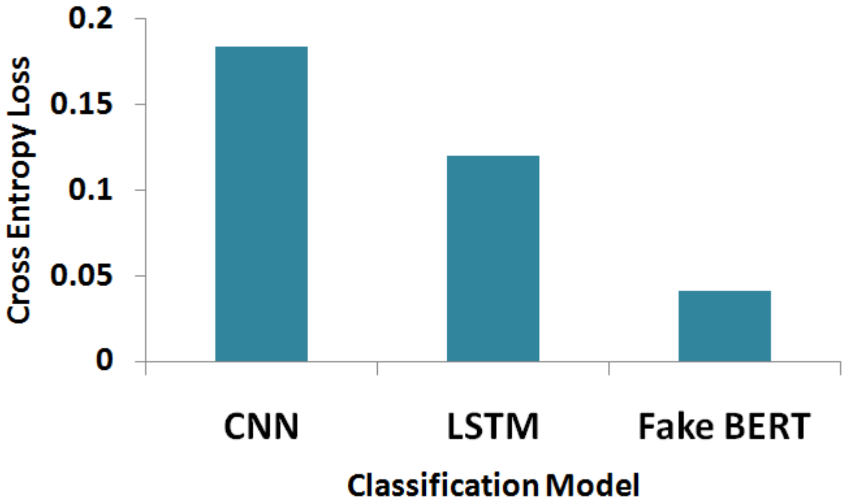

FIGURE 3.11: Cross Entropy Loss with CNN,LSTM,and FakeBERT

deep learning-based approach built on the top of BERT. In the deep investigation, the GloVe-based deep approach was found suitable with LSTM and CNN as a classifier; the improved classification results were found with an accuracy of 92.70% and 97.55% respectively with ten epochs. The respective confusion matrix showed with the help of table 3.31. Respective confusion matrices showed with the help of table 3.32 and 3.33. It was found in the investigation that the BERT approach provided more precise results in fake news classification.
Several experiments conducted with the optimal hyperparameters to validate the designed BERT-based deep learning model (FakeBERT). The proposed investigation found that the model achieved more accurate results with an accuracy of 98.90%. A respective confusion matrix showed with the help of table 3.34. In the proposed approach, the selection of hyperparameters showed in table 3.24. From figure 3.8, the accuracy and cross-entropy loss of our implemented CNN model were examined with a real-world fake news dataset. As seen from figure 3.10, the training loss decayed more quickly with the BERT-based model than the previous word embedding model (like GloVe, word2Vec etc.) From figure 3.6 and table 3.21, the architecture of our implemented BERT-based model (FakeBERT) was presented. Table 3.36 displayed the implemented FakeBERT model's accuracy with 98.90% using the test set. As investigated above, the pre-trained embedding-based models consistently outperformed with a significant margin of improvement. The training loss of the BERT approach decayed comparatively fast and without any inconsistencies. It showed clearly from figure 3.10 that cross-entropy loss reduced fast using the FakeBERT model. More accurate results were achieved with the proposed model than other implemented models with minimal data losses. To validate the recommended

model's performance, two more evaluations parameters (FPR and FNR) were considered. Results tabulated in table 3.37- 3.36. In these results, it was clear that with the proposed model (FakeBERT), both FPR and FNR were minimal, with a value of 1.60% and 0.59%, respectively. With other classification models, the values of FPR and FNR were high.
Using bidirectional pre-trained word embedding (BERT) leads to faster training of model and lower cross-entropy loss. Consistently in the classification tasks, precision and recall improved pre-trained word embedding was used. Table 3.36 showed the results using both machine learning and deep learning models. It was demonstrated that the designed model (FakeBERT) achieved excellent results compared to existing benchmarks or different classification models. From table 3.35, the proposed method's comparative analysis comprehended with the existing classification results using the Kaggle real-world fake news dataset. It was a precise observation that the highest classification accuracy reported with an accuracy of 93.50%. Table 3.36 demonstrated clearly that the recommended model gave comparatively more correct classification results and better performances (testing accuracy, FPR, FNR, Cross-entropy loss). Cross-Entropy loss was also significantly less using BERT as a training model (more details refer to figure 3.11). Using our BERT-based in-depth convolutional approach (FakeBERT), an accuracy of 98.90% was achieved compared to 98.36% with GloVe.

3.3 Summary

In this research, the design and classification performance of the proposed FNDNet and DeepFakE presented. Classification results demonstrated that the deep learning model (FNDNet) provided more correct classification results for predicting fake news with an accuracy of 98.36%. Different performance evaluation parameters (accuracy, true negative rate, recall, precision etc.) included validating the classification results. With the proposed model (FNDNet), there were significantly fewer chances for inaccurate classification, having a very less FPR (0.59%) and a high TNR (99.41%). The cross-entropy rate was also less with the proposed model. Results strongly encouraged to use the proposed model in the area of fake news classification. This research also demonstrated the proposed model's performance (FakeBERT-a BERT-based in-depth convolutional approach) for fake news detection. The proposed model combined BERT and three parallel blocks of 1d-CNN with different kernel-sized convolutional layers with varying filters for better learning. The proposed model built on top of a bidirectional transformer encoder-based pre-trained word embedding model (BERT). Classification results demonstrate that FakeBERT provided more accurate results with an accuracy of 98.90%. The accuracy of

FakeBERT was better than the existing detection systems with real-world fake news dataset: Fake-News.

Chapter 4

Tensor decomposition-based deep learning models for fake news detection *

*The results presented in this chapter are published in: 1. **Kaliyar, R. K.**, Goswami, A., & Narang, P. (2020). DeepFakE: improving fake news detection using tensor decomposition-based deep neural network. The Journal of Supercomputing, 1-23.https://doi.org/10.1007/s11227-020-03294-y.
(Springer Nature, SCI Impact Factor: 2.469).
2. **Kaliyar, R. K.**, Goswami, A., & Narang, P. (2021) EchoFakeD: improving fake news detection in social media with an efficient deep neural network. Neural Computing and Applications, 1-17.https://doi.org/10.1007/s00521-020-05611-1.**(Springer Nature, SCI Impact Factor: 4.774).**

This chapter, the detection of fake news with DeepFakE and EchoFakeD (the proposed deep learning approaches), has been presented.

4.1 DeepFakE

This section presented the research work using the proposed model (DeepFakE) for the detection of fake news.

4.1.1 Introduction

This chapter presented the research work with the proposed model (DeepFakE) for fake news identification. In this research, the news-user engagement (relation between user-profiles and the available news articles) was captured and combined with user-community data (information about the users with having the same perception about a news article) to form a 3-mode (content, context, and user-community) tensor. A tensor was a multidimensional array that gave a higher dimensional generalization (Rabanser et al. 2017) of matrices. A coupled matrix-tensor factorization resulted in an underlying representation of both news content and social context. Experiments conducted using machine learning as well as deep learning-based models. XG-Boost and proposed model (DeepFakE) employed for modelling the aforementioned combined representation for fake news detection. The extra hidden layers enable the composition of features from lower layers, potentially modelling the data. A higher number of hidden layers in the neural network increases weights and help make a higher-order decision boundary. With more hidden layer, there were more chances to approach the end goal quickly. The proposed model was designed containing four hidden layers with the selection of optimal hyperparameters. The proposed model (DeepFakE) achieved better results compared to existing methods with a validation accuracy of 85.86 % with BuzzFeed and 88.64% with the PolitiFact dataset. The contributions of this research were:

- Modeling a deep neural network (DeepFakE) on both news content and social context to obtain more accurate results.
- Demonstrating the classification results on a real-world dataset (BuzzFeed & PolitiFact).

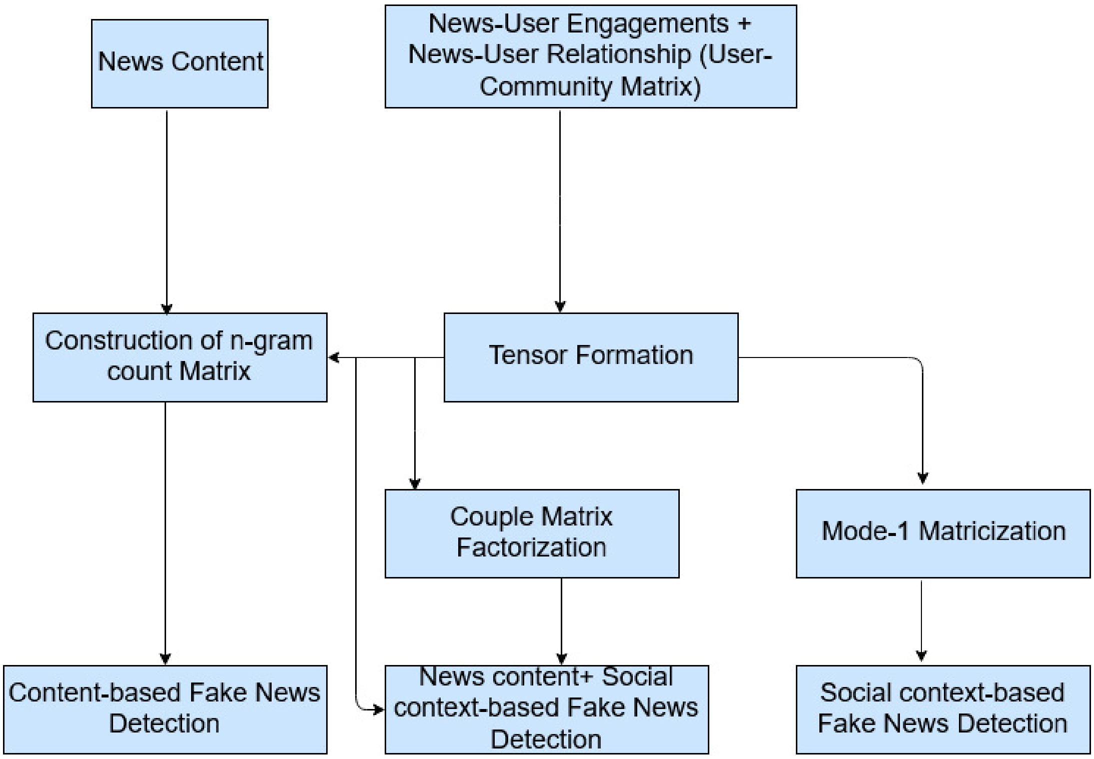

FIGURE 4.1: Proposed Methodology

4.1.2 Experimental setup and methodology

In this section, various experiments and methods presented that were utilized to achieve the research objectives and the results of the experiments. The methodology adopted for fake news detection showed in Fig. 4.1.

Mathematical Representations

Scalar is denoted by small letter (e.g. a), matrix is denoted by capital letter (e.g. A) and tensor is denoted by boldfaced capital letter (e.g. $\boldsymbol{A}$).

Construction of n-gram count matrix

The n-gram count matrix gives a representation of the textual content of the news article. This matrix is denoted by N and has dimensions of $(n \times v)$, where n is the total number of news articles, and v denotes the number of words. Every element in the matrix represents the count of n-gram words in a particular news article.

Construction of news-user engagement matrix

The matrix represents the user response to a news article in terms of sharing. This matrix is denoted by U and has dimensions of $(n \times u)$, where n is the total number of news articles, and u denotes the total users on social media. The elements in this matrix represent the number of times a particular user shared a news article on the social media platform.

Construction of user-community matrix

The user-user relationship given in the dataset exploited to construct the user-community matrix. Meaningful communities extracted from the user network by the Clauset-Newman-Moore algorithm, which was a computationally resource-efficient algorithm (Clauset et al. 2004). Every step of this algorithm merges two communities that contributed the most to the global modularity. The detected communities (Zhang et al. 2016) from the user network thus obtained. The User-community matrix constructed from these detected communities. C denotes the user-community matrix. It has dimensions of $(u \times c)$, where u was the number of users, and c was the number of detected communities. It was a binary matrix where only those elements belonging to a particular community were 1.

Tensor Formation

This research intended to capture the social context-based features present in a news article. These features obtained by representing a news article in news-user engagement and its propagation with user communities' help. Thus, following one step further, tensor modelling of the fake news classification problem was recommended. The latent relation of news articles and the contextual connection between users was captured. Standard factorization methods (Gupta et al. 2018) have shown limited effectiveness due to their unsupervised nature. A news article was presented as a 3-mode tensor (Khatri & Rao 1968) of the structure - news, user, and user cohort (echo chamber). Thus, a tensor decomposition-based method was proposed to encode the news article in a latent representation (Rabanser et al. 2017) preserving the user cohort's structure to produce effective results. The proposed architecture showed in figure 4.2. A tensor was formed, as shown in Equation 4.1.

$$T_{ijk} = U_{ij} * C_{jk} \tag{4.1}$$

The elements in the tensor formed thus depict how a news article propagated in the echo-chambers. The tensor gave a latent representation of the news article based on its social context.

Mode-i Matricization of tensor

Matricization operation re-orders a tensor into a matrix (Rabanser et al. 2017). A mode-i tensor $\boldsymbol{T}$ can be represented such that $\boldsymbol{T} \in R^{I_1 \times I_2 \ldots \times I_i}$. The mode-i matricization of the tensor $\boldsymbol{T}$ obtained from Equation (4.2).

$$X_i \in R^{I_i \times (\Pi_{n \neq i}^{3} I_n)} \tag{4.2}$$

The matrix X_1 was mode-1 matricization of the tensor and has dimension $n \times (u * c)$.

CMTF

The collective data (news content and social context together) was fused by employing Coupled Matrix-Tensor Factorization (CMTF) (Torlay et al. 2017) as presented in (Acar et al. 2011), (Chen & Guestrin 2016). This technique solved the optimization objective, as stated in Equation (4.3).

$$min \ \frac{1}{2} \parallel \boldsymbol{T} - [\![T_1, T_2, T_3]\!] \parallel_F^2 + \frac{1}{2} \parallel N - [\![N_1, N_2]\!] \parallel_F^2 \tag{4.3}$$

In the above equation, $\boldsymbol{T}$ was the tensor with news, user and community information. $[\![T_1, T_2, T_3]\!]$ represents the Kruskal operation on matrices T_1, T_2 and T_3, such that $T_1 \in R^{I_1 \times R}$, $T_2 \in R^{I_2 \times R}$ and $T_3 \in R^{I_3 \times R}$. These matrices were obtained by factorizing the tensor using the R-component PARAFAC procedure (Harshman et al. 1970). The matrix N was the news content matrix and N_1 and N_2 were the R factor matrices obtained after matrix factorization (Lee & Seung 2001) of N, where $N_1 \in R^{n \times R}$ and $N_2 \in R^{v \times R}$. Equation (4.3) can be rewritten as shown in Equation (4.4).

$$min \ \frac{1}{2} f_1 + \frac{1}{2} f_2 \tag{4.4}$$

The above optimization problem was solved by computing gradients of the components f_1 and f_2 with respect to factors. The computation of gradients showed in Equations (4.5)-(4.7).

$$\frac{\partial f_1}{\partial T_i} = (Z_i \; - \; X_i)\, T_i^{-i} \tag{4.5}$$

$$\frac{\partial f_2}{\partial N_1} = -NN_2 + N_1^{-1} N_2^T N_2 \tag{4.6}$$

$$\frac{\partial f_2}{\partial N_2} = -N^T N_1 + N_2 N_1^T N_1 \tag{4.7}$$

where,

$$Z = [\![T_1, T_2, T_3]\!] \tag{4.8}$$

$$Z_1 = T_1\,(\,T_3 \,\odot\, T_2\,)^T \tag{4.9}$$

$$Z_2 = T_2\,(\,T_3 \,\odot\, T_1\,)^T \tag{4.10}$$

$$Z_3 = T_3\,(\,T_2 \,\odot\, T_1\,)^T \tag{4.11}$$

$$T^{-i} = T^I \odot \; ... \; T^{i+1} \odot \; T^{i-1} \odot \; .... \; \odot T^1 \tag{4.12}$$

The symbol $\odot$ in Equations (4.9-4.12) represented Khatri-Rao product (Khatri & Rao 1968). X_i in Equation (4.5) was mode-i matricization of tensor T. The final gradient matrix was formed by the concatenation of vectorized partial derivatives with respect to factor matrices. The final gradient to be obtained expressed in Equation (4.13).

$$\nabla_f = \begin{bmatrix} vec(\frac{\partial f_1}{\partial T_1}) \\ \\ vec(\frac{\partial f_1}{\partial T_2}) \\ \\ vec(\frac{\partial f_1}{\partial T_3}) \\ \\ vec(\frac{\partial f_2}{\partial N_1}) \\ \\ vec(\frac{\partial f_2}{\partial N_2}) \end{bmatrix} \tag{4.13}$$

Conjugate gradient algorithm used for minimization of objective function. The factor matrices obtained after optimization are a lower dimensional representation of the tensor which denotes news, user and community information. The first mode factor obtained after factorization used as a feature for classification.

4.1.3 Classification

Experiments were conducted to classify news articles utilizing both the content of news and social context-based characteristics separately and a combination of both.

- News Content-based Classification: The n-gram count matrix representing only the textual content of news used for classification.
- Social Context-based Classification: The mode-1 matricized tensor described news interaction with users, and it used for classification based on the social context.
- News Content + Social Context-based Classification: The n-gram count matrix concatenated with the first mode factor. The resulting matrix used for classification based on news content and social context.

Machine Learning Approach

Experiments were conducted using a decision tree-based model (XGBoost). The detailed description of XGBoost described below:

- **XGBoost Algorithm:-** XGBoost (Chen & Guestrin 2016; Torlay et al. 2017) is a fastest ensemble model that utilizes the concept of gradient boosting (Chen & Guestrin 2016; Zhang & Ghorbani 2020). XGBoost and Gradient Boosting Machines (GBM's) both can apply to the standard boosting weak learners (such as CART's (Moreno et al. 2001)) utilizing the gradient descent architecture. XGBoost enhances the base GBM structure through systems optimization and algorithmic enhancements. XGBoost handles the inefficiency of possible splits during feature selection by looking at the distribution of features (Patidar, Sharma, et al. 2011) across all data points in a leaf and utilizing it to reduce the search space of possible feature splits.

Deep Learning Approach:- The Proposed Neural Network (DeepFakE)

In this research, a multi-layer deep neural network was designed (Jain et al. 1996; Zurada 1992) for identifying fake news. The choices of a number of dense layers, dropout, the selection of activation function and loss function to make the proposed detection model more efficient and deep optimized were as follows:

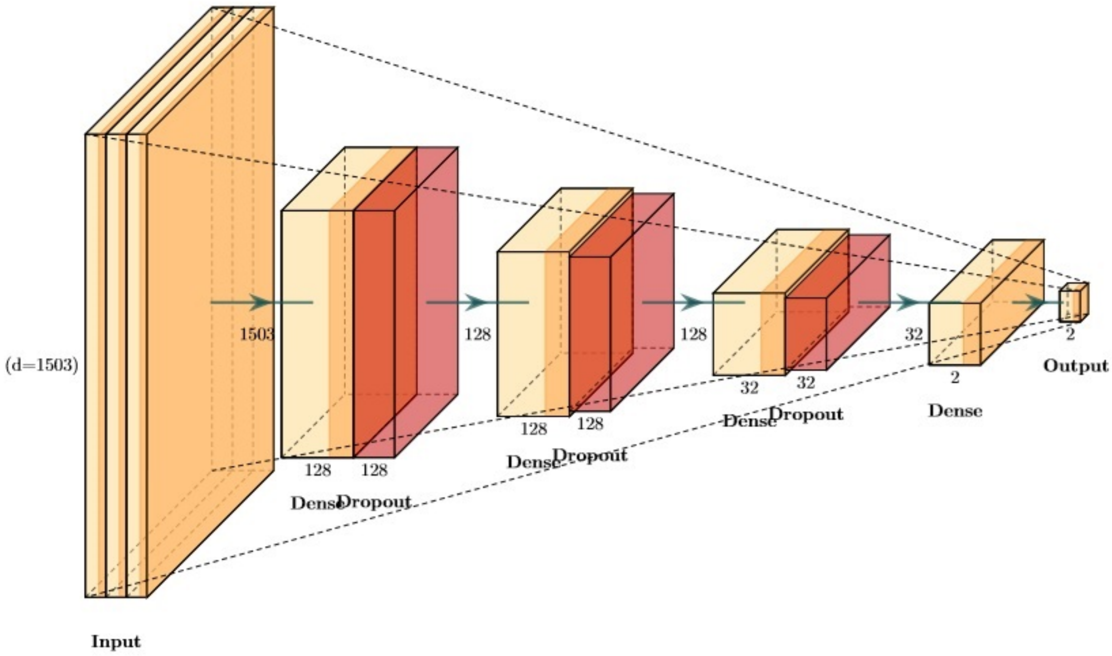

FIGURE 4.2: Architecture of DeepFakE

TABLE 4.1: Layered Architecture of the proposed Deep Neural Network-based model (DeepFakE)

Layer (type)	Input size	Output size
Dense	1503	128
Dropout	128	128
Dense	128	128
Dropout	128	128
Dense	128	32
Dropout	32	32
Dense	32	2

- **Dense layer:** In this research, a deep neural network (DNN) with four hidden layers designed. This network took the obtained features as input and classified the test samples into either one of the categories: fake or real. The addition of hidden layers in the neural network helped to improve the model, but only up to a certain point. Further expansion of layers can harm the model's performance (it depends on the problem's complexity). Researchers (Chen et al. 2014; Patidar, Sharma, et al. 2011; Wang et al. 2018; Zhong et al. 2019) have primarily used one or two dense layers before the final softmax layer. With the proposed deep approach, four dense layers were taken with hyperparameters optimization.

- **Dropout:** Dropout as a regularization technique (Srivastava et al. 2014; Wager et al. 2013; Wu & Gu 2015) which aims to reduce the complexity of any model with the end goal of preventing over-fitting (Vasudevan et al. 2019). The application of dropout at each layer of the network showed promising results. One can understand the dropout concept with an example; the dropout rate is set to 10%, meaning one in 10 inputs will be randomly excluded from each update cycle. In this research, the value of dropout was taken as 0.2.

- **Activation Function:** ReLU (Rectified Linear Unit) was selected as activation function (Li & Yuan 2017). It was the most commonly used activation function in Deep Learning due to efficient results. It was computationally efficient than sigmoid or Tanh and solved the vanishing gradient problem. The equation of ReLU can be written as:

$$\sigma = max(0, z) \tag{4.14}$$

- **Loss Function (L):** Cross-entropy loss estimates the model's performance, whose output lies with a probability value between 0 and 1. In binary classification, cross-entropy can be calculated as:

$$L = -(ylog(p) + (1-y)log(1-p)) \tag{4.15}$$

  If M>2 (i.e. multi-class classification), a separate loss for each class label is calculated label per observation and sum the result.

$$-\sum_{c=1}^{M} y_{o,c} log\left(p_{o,c}\right) \tag{4.16}$$

  Here, M - the number of classes
  log - the natural *log*

TABLE 4.2: Description of FakeNewsNet Dataset

News Source	Number of News Articles	Number of Fake News Articles	Number of Users
BuzzFeed	182	91	15257
PolitiFact	240	120	23865

y - binary indicator (0 or 1), c-class, o-observation
p - predicted probability

In figure 4.2 and table 4.1, the layered architecture of the proposed multi-layer-based DNN model (DeepFakE) showed. In this network, a total of 1503 input nodes and four hidden layers were considered. The first hidden layer has 128 nodes and a dropout of 0.2. The second layer also has 128 hidden nodes with no dropout. The third layer has 128 hidden nodes and a dropout of 0.2. The fourth layer has 32 hidden nodes and a dropout of 0.2. The final output layer has two nodes and an activation function as SoftMax. The experiments carried out using the NVIDIA DGX-1 V100 machine. The machine equipped with 40,600 CUDA cores, 5,120 tensor cores, 128 GB RAM and 1,000 TFLOPS speed.

4.1.4 Experimental results

Dataset

The proposed method examined on BuzzFeed and PolitiFact dataset from the FakeNewsNet Dataset [†]. The dataset contains the following information:

- Real and fake news content: Contains news articles with attributes such as news id, title, text, URL, authors and source.
- News and user engagement: Specifies the number of times a user has shared a news article.
- User-user relationship: Specifies the user network on social media.

A brief description of FakeNewsNet dataset has given in Table 4.2.

[†] The dataset can be downloaded from https://www.kaggle.com/mdepak/fakenewsnet

TABLE 4.3: Hyperparameters for the proposed Deep Neural Network-based model

Hyperparameter	Description or Value
No. of Dense layers	4
No. of Hidden nodes	128,128,32,2
Dropout rate	0.2
Activation function	ReLU
Loss function	Binary cross-entropy
Output layer	Softmax
Number of epochs	20
Batch-size	32
Learning rate	0.1
Optimizer	Adam

TABLE 4.4: Dimensions of features

Features	Dimensions
n-gram count matrix (N)	(182 x 1500)
News-user engagement matrix (U)	(182 x 15257)
User-community matrix (C)	(15257 x 81)
Tensor $\boldsymbol{T}$	(182 x 15257 x 81)
Mode-1 matricized tensor (X_1)	(182 x (15257x81))
Combined content + context matrix	(182 x 1503)

Hyperparameter settings (refer table 4.3 for more details)

- Feature Extraction:
  Sklearn library in Python used to construct the n-gram count matrix. The number of words in the vocabulary was limited to 1500. The number of communities obtained after the Clauset-Newman-Moore algorithm is 81. Table 4.2 showed that the number of news articles was 182, and the total number of users was 15257 for the BuzzFeed dataset. The number of news articles was 240, and the total number of users was 23865 for the PolitiFact dataset. The dimensions of all the matrices used as features for the classification task have given in Table 4.4.

- Deep Neural Network:
  The DNN with four layers with 128, 128, 32 and 2 hidden nodes respectively designed.

TABLE 4.5: Comparison benchmark results using FakeNewsNet Dataset (BuzzFeed)

Authors	Precision (%)	Recall (%)	F1-Score(%)
(Castillo et al. 2011)	73.50	78.30	75.60
RST - (Castillo et al. 2011)	79.50	78.40	78.90
CITDetect- (Gupta et al. 2018)	65.70	100.00	79.20
CIMTDetect- (Gupta et al. 2018)	72.90	92.30	81.30
CLASS-CP- (Papanastasiou et al. 2019)	85.20	83.00	83.50
DeepFakE-the proposed model	**83.33**	**86.96**	**85.11**

TABLE 4.6: Comparison benchmark results using FakeNewsNet Dataset (PolitiFact)

Authors	Precision (%)	Recall (%)	F1-Score(%)
(Castillo et al. 2011)	77.70	79.10	78.30
RST - (Castillo et al. 2011)	82.30	79.20	79.30
CITDetect- (Gupta et al. 2018)	67.90	97.50	79.10
CIMTDetect- (Gupta et al. 2018)	80.30	84.20	81.80
DeepFakE-the proposed model	**82.10**	**84.60**	**84.04**

ReLU with $\alpha = 0.1$ used as activation function. The weights initialized from a normal distribution presented in (He et al. 2015). Adam optimizer used for optimizing the designed DNN. DNN has trained for 20 epochs. The dropout regularization method employed to avoid over-fitting.

Performance Parameters

To assess the proposed model's performance, precision, recall, F1-Score, confusion matrices, and validation accuracy were used as evaluation matrices.

4.1.5 Experiments

The following approaches implemented for fake news detection:

- News content + XGBoost classifier (Natekin & Knoll 2013): In the proposed approach, the input feature matrix was the n-gram count matrix N. This matrix represented only the textual content of the news article.

- Social context + XGBoost classifier: The features for this approach were only social context-based. The matrix (X_1) was obtained after the mode-1 matricization of the tensor, which used as an input feature matrix.

TABLE 4.7: Confusion matrix for news content-based classification using XGBoost classifier (BuzzFeed)

	Predicted negative	**Predicted positive**
Actual negative	19 (TN)	8 (FP)
Actual positive	8 (FN)	20 (TP)

TABLE 4.8: Confusion matrix for social context-based classification using XGBoost classifier (BuzzFeed)

	Predicted negative	**Predicted positive**
Actual negative	26 (TN)	2 (FP)
Actual positive	7 (FN)	20 (TP)

- News content and social context + XGBoost Classifier: This was a combined approach using content and context information. The input feature matrix obtained after concatenating the n-gram count matrix with the mode-1 factor matrix after coupled matrix-tensor factorization.
- News content + social context + DNN: The combined news content and social context features used for classification using DNN. The results of all these approaches compared.

Results

The classification results tabulated in Tables 4.11 & 4.12. Precision, recall, F1-Score, and accuracy calculated from the confusion matrix and used to evaluate classification results. Table 4.6 showed that combining news content and social context-based features gives better results by employing DNN compared to other approaches. The performance of XGBoost classifier summarized with a confusion matrix. The elements of the matrix gave the count of correct and incorrect classifications. The confusion matrices for the machine learning and the proposed approach showed in tables 4.7-4.10.

TABLE 4.9: Confusion matrix for news content and social context-based classification using XGBoost classifier (BuzzFeed)

	Predicted negative	**Predicted positive**
Actual negative	19 (TN)	8 (FP)
Actual positive	8 (FN)	20 (TP)

TABLE 4.10: Confusion matrix for news content and social context-based classification using DeepFakE (BuzzFeed)

	Predicted negative	**Predicted positive**
Actual negative	19 (TN)	4 (FP)
Actual positive	3 (FN)	20 (TP)

TABLE 4.11: Classification Results using BuzzFeed

Approach	**Precision**	**Recall**	**F1-Score**	**Validation Accuracy**
News Content + XGBoost	0.714	0.714	0.714	0.709
Social Context + XGBoost	0.74	0.909	0.815	0.836
News Content and Social Context + XGBoost	0.714	0.714	0.714	0.709
News Content and Social Context + DNN (DeepFakE)	0.8333	0.8696	0.8511	0.8649

TABLE 4.12: Classification Results using PolitiFact

Approach	**Precision**	**Recall**	**F1-Score**	**Validation Accuracy**
News Content + XGBoost	0.7454	0.7720	0.7437	0.7880
Social Context + XGBoost	0.7868	0.9135	0.815	0.8454
News Content and Social Context + XGBoost	0.8034	0.9520	0.8714	0.8670
News Content and Social Context + DNN (DeepFakE)	0.8210	0.8460	0.8404	0.8864

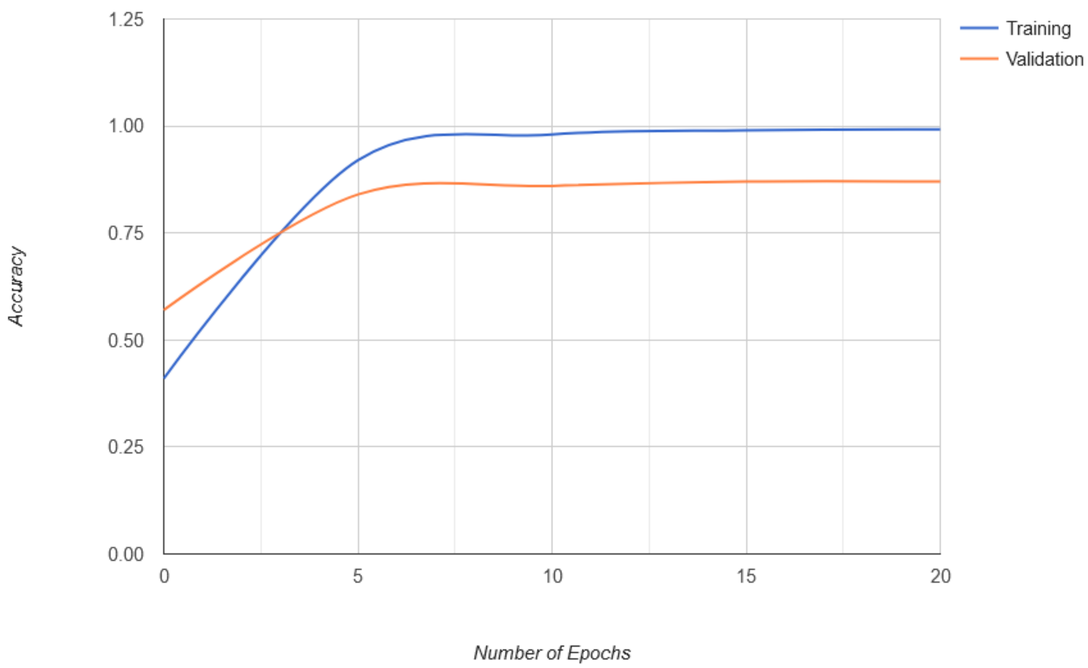

FIGURE 4.3: DNN accuracy and Cross-entropy loss curves for combined content and context-based classification

Classification only news content and combined news content with social context-based data with XGBoost classifier classified the news articles misclassified in a large. The confusion matrices observed that social context-based classification performs the best among all three approaches using machine learning. The DNN accuracy and cross-entropy loss with training accuracy versus the number of epochs revealed that combining news content and social context showed in figures 4.3 & 4.4. The curves that the model learned well and did not over-fit the training data observed.

The proposed method outperformed existing fake news detection methods. It considered the textual attributes of news articles and news articles' interaction with users on social media. A news article's social context was a latent feature extracted from the tensor representing news-user engagement and user-user relationship. The features obtained after the factorization method capturing the news, user, and community inter-dependencies and thus represent the news article.

From tables 4.11 & 4.12, the value of different performance parameters showed. With the proposed model, the precision value was 0.8333 and 0.8210, respectively, with BuzzFeed and PolitiFact dataset. High precision relates to the low false-positive rate. The proposed model achieved 0.8333 precision with the BuzzFeed dataset, which showed better results. The value

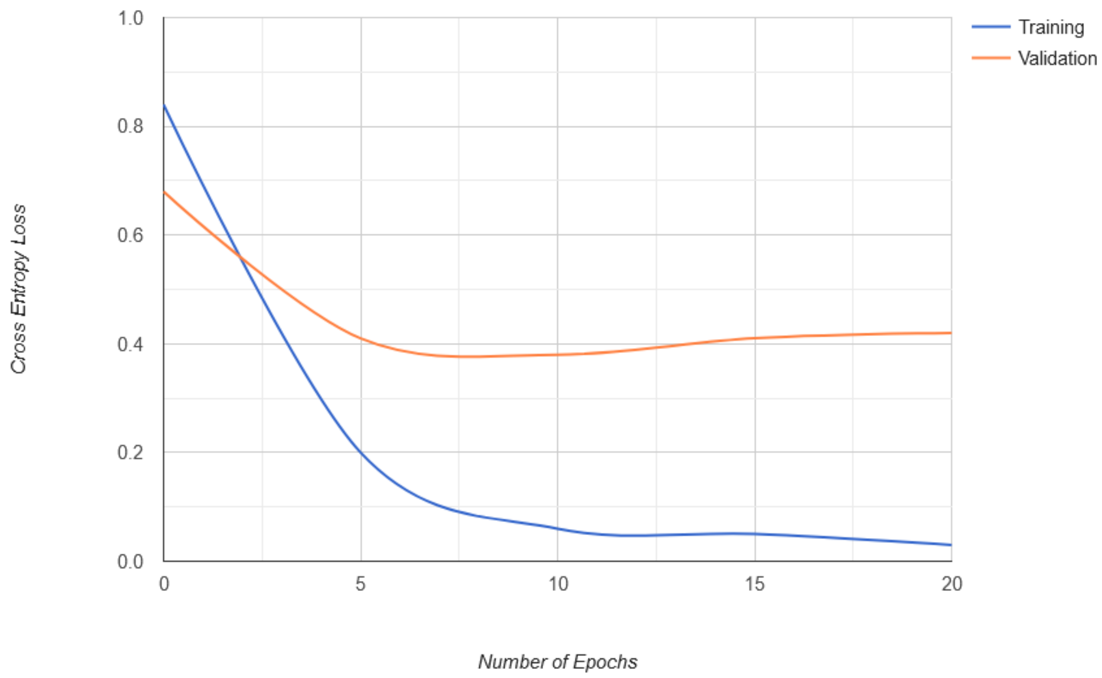

FIGURE 4.4: DNN accuracy and Cross-entropy loss curves for combined content and context-based classification

of recalls was 0.8696 and 0.8460, respectively, using both datasets. The recall value was above 0.50, which showed better results in terms of F1-Score, which takes both false positives and false negatives into account. Using the proposed model, the value of F1-Score was 0.8511 and 0.8404, respectively, with BuzzFeed and PolitiFact dataset. The values of F1-Score was higher means better classification results. With the proposed model, a training accuracy of 0.9904 and 0.9931 achieved and a validation accuracy of 0.8649 and 0.8864 achieved using the BuzzFeed and PolitiFact dataset. More accurate results were achieved utilizing both news content and social context-based features. The proposed DNN further improved the classification performance compared to existing traditional methods. From tables, 4.5 & 4.6, the performance of the existing benchmark models was shown with the proposed model using the context-related fake news dataset. Results demonstrated the proposed method's superior performance (DeepFakE) for fake news detection compared to other existing benchmarks.

4.2 EchoFakeD

This section presented the research work using the proposed model (EchoFakeD) for fake news identification combining content, context, and user-communities.

4.2.1 Introduction

This research presented the work with an effective deep learning model utilizing content and context-related features. In one of the present study, Gupta et al. (Gupta et al. 2018) investigated the problem of fake news with the user and community-level information using the tensor factorization approach. In their process, two methods (CITDetect and CIMTDetect) proposed using both content and user-community information as a combination. They experimented with a traditional machine learning classifier (SVM) and achieved an accuracy of 81.30% and 81.80%, respectively, with both real-world fake new datasets: BuzzFeed and PolitiFact. In their method, information about the dimensional features was vague. Deep learning techniques are capable of learning more abstract representation of data as the network grown deeper; thus, the model automatically extracts features and yields higher accuracy results. Keeping these points in mind, the proposed model was one step ahead of the existing approach. In this process, for textual modality and effective detection, we performed extensive feature set-based studies to classify fake news. In this research, the user's engagement with the news articles is captured and fused with user-community interaction to form a 3-mode tensor (content, social context, and user-community information). This tensor handled multi-relational data (Rabanser et al. 2017) and provided a higher dimensional generalization of matrices. Tensor factorization decomposes the higher-order tensor into low-rank tensors. The resulting low-rank tensors (Gupta et al. 2018; Rabanser et al. 2017) captured the complex relations between the objects representing the help of models of the tensor. We achieved the dimension of a combined matrix (content-context information) is 182x1503, in which many news stories were 182, and the size of the input word embedding was 1503. Standard factorization methods (Hosseinimotlagh & Papalexakis 2018) showed limited effectiveness due to their unsupervised environment. In the coupled matrix-tensor factorization method (also known as CP-decomposition), we utilized the standard factorization method to decompose the matrix. In labelled data, the class information could help the factorization process identify fake news better. Then, the proposed network used for modelling this combined representation of fake news information. A thin deep network (Hosseinimotlagh & Papalexakis 2018; Paisitkriangkrai et al. 2015; Zagoruyko & Komodakis 2015) with two or three hidden layers outperformed all traditional methods by a significant margin (2-4%) on handcrafted features. The architecture worked fine on both small and large datasets and reduced classification error. Deep neural networks eliminated feature engineering and handled high dimensional datasets with millions of parameters that pass through non-linear function. Therefore, five dense layers to make our model deep in nature were considered. After increasing more hidden layers with the proposed neural network, it was likely to over-fit the model and, in turn, depreciate accuracy on the test data. Using user community-based features with news content as a large dimensional tensor, the optimal results were achieved with a neural network having five hidden layers. The proposed deep learning model (EchoFakeD) employed content and context-based information to validate the classification results. The designed model outperformed as compared to the existing baselines and obtained an accuracy of 92.30%. The main contributions of this research were:

- Presented extensive feature set-based studies for the classification of fake news
- Designed an efficient Deep Neural Network consolidating the content level features of news articles with user's social engagement (Echo-chamber infused) to produce significant results
- Implemented a tensor factorization-based approach with content as well as context-based information.
- Utilized an Echo-chamber infused 3-mode Tensor for higher-dimensional generalization

4.2.2 Experimental setup and methodology

In this section, various experiments and techniques presented that were utilized to achieve the research objectives and the results of the experiments. The complete methodology adopted in our research presented in details. The designed method with tensor factorization approach displayed in Fig. 4.5.

Mathematical Representations

In this research, scalar is represented by (e.g. a), matrix is represented by (e.g. A) and tensor is represented by boldfaced capital letter (e.g. $\boldsymbol{A}$).

Construction of count matrix using textual content

The matrix is represented by N having the dimensions of $(n \times v)$, where n is the total number of news articles and v is the number of words. This matrix is used to count the sequences of words in a news article.

Construction of news-user engagement matrix

In this matrix, the response of the user to a particular news article has been shown. The matrix is denoted by U having the dimensions of $(n \times u)$, where n is the total number of news articles and u is the number of users. This matrix is used to represent the counting of news articles shared by any particular user on social media.

Development of user-community matrix

A heuristic method for community detection in large scale networks reported by Blondel et al. (Blondel et al. 2008), which employed an agglomeration multi-step process during its execution. Wakita et al. (Wakita & Tsurumi 2007) also discussed community detection for small networks. In this research, the user-user relationship considered in the available information to build the user-community matrix. The method suggested by Clauset et al. (Clauset et al. 2004) used for the fast identification of communities. This algorithm was computationally resource-efficient and designed for large and complex networks. It associated each node of the network with a community. The proposed algorithms' computational complexity analyzed without compromising the model's performance. In this algorithm, the main step was combining two communities that mainly contribute to global modularity. Clauset et al. proposed a matrix M to store a modularity gain (Clauset et al. 2004) by the union of two communities C_a and C_b when the communities (Wakita & Tsurumi 2007) were connected. The elements M_{ab} of Matrix M were initialized by equation 4.18. Here d was an element that stores the sum of degrees of the modes in a network that belongs to a particular community (refer equation 4.17). The User-community matrix represented C having the dimensions of $(u \times c)$, where u was the number of users, and c as the number of detected communities. According to Clauset, the matrix M for each union was updated until we got no more modularity gain. Clauset also defined the rules for updating (refer equation 4.19) the whole matrix M concerning connected communities combined with

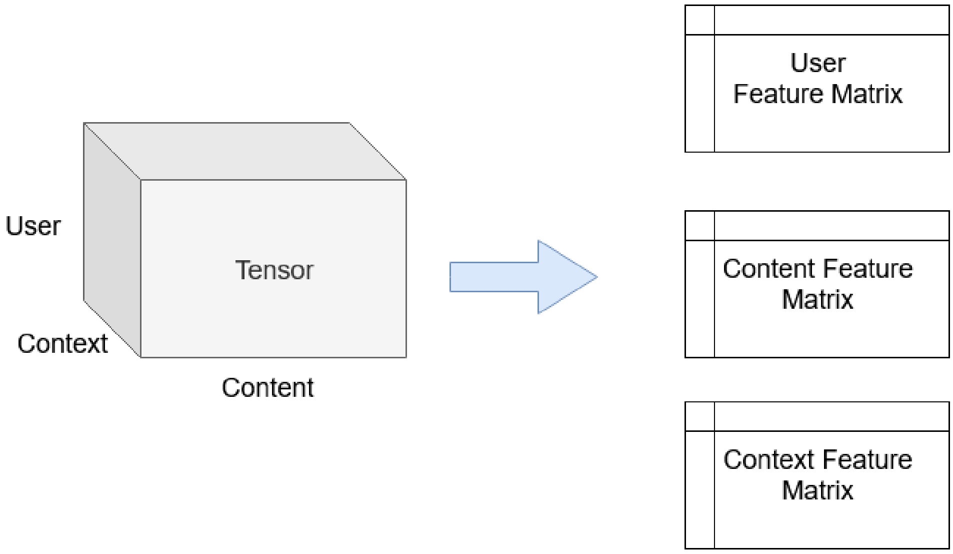

FIGURE 4.5: Tensor Decomposition Approach

other communities.

$$M_{ab} = \begin{cases} \frac{1}{2m} - \frac{d_a d_b}{(2m)^2}, & if\ C_a\ and\ C_b\ are\ connected \\ 0, & otherwise \end{cases} \tag{4.17}$$

$$Degree\ of\ vertex\ a = d_a = \sum_i k_i,\ v_i \varepsilon\ C_a \tag{4.18}$$

$$M'_{ac} = \begin{Bmatrix} M_{ac} + M_{bc}, & if\ C_c\ is\ connected\ to\ C_a\ and\ C_b \\ M_{bc} - 2\frac{d_a}{(2m)}\frac{d_c}{(2m)}, & if\ C_c\ is\ connected\ to\ C_b\ but\ not\ to\ C_a \\ M_{ac} - 2\frac{d_b}{(2m)}\frac{d_c}{(2m)}, & if\ C_c\ is\ connected\ to\ C_a\ but\ not\ to\ C_b \end{Bmatrix} \tag{4.19}$$

Here, m = number of edges in the network, k=degree vector, v=a vertex in the network, d=a vector which stores the sum of degree of nodes

Formation of tensor

A tensor formed as shown in Equation (4.20). The representation of a 3-mode tensor showed in figure 4.5 which consists of a combination of different feature matrices.

$$T_{ijk} = U_{ij} * C_{jk} \tag{4.20}$$

Re-ordering of tensor using matricization

A tensor can be re-ordered into a matrix using matricization operation (Rabanser et al. 2017). We can represent a mode-i tensor $\boldsymbol{T}$ such that $\boldsymbol{T} \in R^{I_1 \times I_2 \ldots \times I_i}$. The mode-i matricization of the tensor $\boldsymbol{T}$ represented by Equation (4.21).

$$X_i \in R^{I_i \times (\Pi_{n \neq i}^{3} I_n)} \tag{4.21}$$

Here, the matrix X_1 represents mode-1 matricization having dimension $n \times (u * c)$. Here, i is defined in the range of [1,3].

A Coupled Matrix-Tensor factorization approach

The combination of both news content and social context fused by employing Coupled Matrix-Tensor Factorization (CMTF) method (Acar et al. 2011; Gupta et al. 2018). This approach solved the optimization, as stated in Equation (4.22).

$$min \; \frac{1}{2} \parallel \boldsymbol{T} - [[T_1, T_2, T_3]] \parallel_F^2 + \frac{1}{2} \parallel N - [[N_1, N_2]] \parallel_F^2 \tag{4.22}$$

In the equation (4.22), $\boldsymbol{T}$ was the 3-mode tensor (news, user and community). $[[T_1, T_2, T_3]]$ represented the Kruskal operation on matrices T_1, T_2 and T_3, such that $T_1 \in R^{I_1 \times R}$, $T_2 \in R^{I_2 \times R}$ and $T_3 \in R^{I_3 \times R}$. These matrices can be obtained by factorizing the tensor using the R-component PARAFAC procedure (Harshman et al. 1970). In the equation, N denotes the news content matrix and N_1 and N_2 are the R-factor matrices obtained after non-negative matrix factorization (Lee & Seung 2001) of N, where $N_1 \in R^{n \times R}$ and $N_2 \in R^{v \times R}$. Equation (4.22) can be re-written as Equation (4.23).

$$min \; \frac{1}{2} f_1 + \frac{1}{2} f_2 \tag{4.23}$$

Optimization problem can be solved by computing gradients of the components f_1 and f_2 with respect to factors and shown with the help of Equations (4.24)-(4.26).

$$\frac{\partial f_1}{\partial T_i} = (Z_i - X_i)\, T_i^{-i} \tag{4.24}$$

$$\frac{\partial f_2}{\partial N_1} = -NN_2 + N_1^{-1} N_2^T N_2 \tag{4.25}$$

$$\frac{\partial f_2}{\partial N_2} = -N^T N_1 + N_2 N_1^T N_1 \tag{4.26}$$

where,

$$Z = [[T_1, T_2, T_3]] \tag{4.27}$$

$$Z_1 = T_1 \,(\, T_3 \odot T_2 \,)^T \tag{4.28}$$

$$Z_2 = T_2 \,(\, T_3 \odot T_1 \,)^T \tag{4.29}$$

$$Z_3 = T_3 \,(\, T_2 \odot T_1 \,)^T \tag{4.30}$$

$$T^{-i} = T^I \odot \ldots T^{i+1} \odot T^{i-1} \odot \ldots . \odot T^1 \tag{4.31}$$

The symbol $\odot$ in Equations (4.28-4.31) represented Khatri-Rao product (Khatri & Rao 1968). The final gradient matrix can be obtained by the concatenation of vectorized partial derivatives with respect to factor matrices as expressed in Equation (4.32).

$$\nabla_f = \begin{bmatrix} vec(\frac{\partial f_1}{\partial T_1}) \\ \\ vec(\frac{\partial f_1}{\partial T_2}) \\ \\ vec(\frac{\partial f_1}{\partial T_3}) \\ \\ vec(\frac{\partial f_2}{\partial N_1}) \\ \\ vec(\frac{\partial f_2}{\partial N_2}) \end{bmatrix} \tag{4.32}$$

Apart from traditional models (Chen & Guestrin 2016; Natekin & Knoll 2013; Torlay et al. 2017) for fake news detection, the proposed neural network was discussed in this section.

EchoFakeD: The proposed Deep Neural Network

This research presented a deep neural network with numerous filters across each layer with some dropout. Five dense layers to make the proposed model more effective and deep were considered. The selection of activation function, loss function, and dropout, which make the proposed model efficient, summarized below:

- **Dense layer:** A dense layer is just a regular layer of neurons in a neural network (Vasudevan et al. 2019; Zhong et al. 2019). The proposed approach took five dense layers to make the model deep in nature to select optimal hyperparameters.

- **Dropout:** In this research, the dropout at each layer of the network was applied. This functionality showed promising results. Experiments were conducted with the value of dropout to be 0.2.

- **Activation Function:** In the proposed deep learning model, ReLU (Rectified Linear Unit) (Li & Yuan 2017) as the activation function was selected. ReLU significantly accelerates the convergence of stochastic gradient descent compared to other activation functions (Li & Yuan 2017). The equation of ReLU can be defined as:

$$\sigma = max(0, z) \tag{4.33}$$

- **Loss Function (L):** Cross-Entropy is widely used as a loss function for binary classification problems. The cross-entropy goes down as the prediction gets more and more accurate. It becomes zero if the prediction is perfect. Cross-entropy loss (M=2) can be defined as:

$$L = -(ylog(p) + (1-y)log(1-p)) \tag{4.34}$$

  If M>2 (i.e. multi-class classification), a separate loss for each class can be calculated as the sum of the result

$$-\sum_{c=1}^{M} y_{o,c} log(p_{o,c}) \tag{4.35}$$

  Here, M - number of classes, c- a class, log - the natural logarithmic function , y - binary indicator (0 or 1), o- an observation, p - predicted probability.

In figure 4.6 and table 4.13, the layered architecture of the proposed deep neural network (EchoFakeD) showed. In this network, the input was based on word-embedding vectors with 1503 nodes. A neural network consists of five dense layers was designed. The first dense layer contains 128 hidden nodes with a dropout of 0.2. The second dense layer contains 128 hidden nodes without dropout. The third dense layer contains 2048 hidden nodes with a dropout of 0.2. The fourth dense layer contains 32 hidden nodes with a dropout of 0.2. The fifth dense layer contains 32 hidden nodes with a dropout of 0.2. The final layer has two nodes with an activation function as SoftMax. The research work carried using the NVIDIA DGX v100 machine, equipped with 40600 CUDA cores, 5120 tensor cores, 128 GB RAM , and 1000 TFLOPS speed.

TABLE 4.13: Layered Architecture of the proposed network-EchoFakeD

Layer	Input (Number of filters)	Output (Number of filters)
Dense layer	1503	128
Dropout layer	128	128
Dense layer	128	128
Dense layer	128	2048
Dropout layer	2048	2048
Dense layer	2048	32
Dropout layer	32	32
Dense layer	32	2

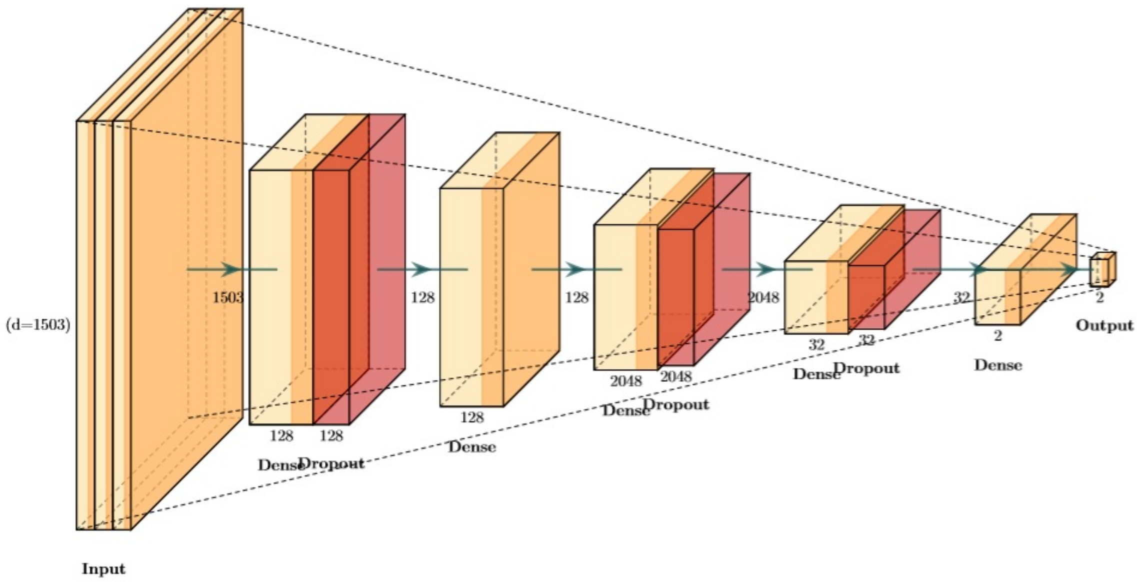

FIGURE 4.6: Architecture of the proposed network-EchoFakeD

4.2.3 Methodology and results

Bench-marking the performance of the designed method, several classifications approach were investigated. These detection methods were based on utilizing the textual news content, user-context, and user-based relations present in the news.

- **Dataset:-**

  Experiments were conducted to verify the performance of the recommended model using the fake news dataset: PolitiFact and BuzzFeed from the FakeNewsNet [‡]. The number of news articles and users in the dataset tabulated in table 4.14 for PolitiFact and BuzzFeed. A total of 145 news articles were taken for training and 37 for testing the model (80:20 ratio). The performance of the proposed model was validated using 37 fake news articles. The valuable information in the dataset as follows:

  - News content: Having the attributes as news-id, URL, title, text, authors, and news source.
  - News-User engagement: It contains the information that how many times a user on social media has shared a news article.
  - User-User engagement: It contains the relationships between the users.

- **Feature Extraction and Hyperparameter Setting**

  - Feature Extraction: In this research, news content, context, and user-community-based features were considered for fake news classification. The Sklearn library used to construct the features matrices. The dimensions of all the matrices (used as input features) showed in table 4.16. A total of 81 communities (featured as Echo chamber) were extracted using the Clauset-Newman-Moore algorithm. (Clauset et al. 2004).
  - Hyperparameter Setting: Hyperparameter (Djidjev 2006; He et al. 2015; MacKay 1999) can be defined as significant elements or variables that are needed during the entire classification process (training and testing the model). In the proposed approach (for more details, refer to table 4.15), hyperparameters were selected before training and optimizing the weights and bias.

[‡] https://www.kaggle.com/mdepak/fakenewsnet

- **Performance Parameters:-**

  To validate the proposed model's performance, different performance parameters considered: precision, recall, accuracy, and confusion matrix as evaluation matrices.

- **Experiments:-**

  To classify the combination of both news content and its social context information, a tensor-based factorization method deployed. The order of classification tasks performed in this research as follows:

  – EchoFakeD (the proposed Deep Neural Network) with news content: For the experiment, the input feature matrix was the count matrix N.

  – EchoFakeD with social context: For an experiment, social context-based has used. Matrix (X_1) obtained after mode-1 matricization operated as an input feature.

  – EchoFakeD with content and social context of news articles: For an experiment, both news content and social context-based features were used for classification. The proposed model obtained satisfactory results with a combination of different features.

- **Experimental Results:-**

  Experiments carried out using the proposed deep learning classifier (EchoFakeD) with different learning paradigms. From figure 4.6, the layered architecture of the proposed model is shown. In this research, a real-world fake news dataset (FakeNewsNet) used for classification.

  Further, experiments were conducted with the proposed model using both contents and news articles' social context. Tables 4.19 and 4.22 showed that the combination of features gives more accurate results by employing a deep neural network. Respective confusion matrices for the deep learning approaches showed with tables 4.17-4.22. The elements of confusion matrices gave the number of correct and incorrect classifications. The proposed model provided better performance than existing benchmarks employing tensor factorization method.

  To validate the achievement of the designed model with the existing methods, several performance parameters (precision, F1-Score, recall etc.) were considered. Complete classification results (using PolitiFact and BuzzFeed dataset) tabulated in tables 4.23 and 4.25. In table 4.25, the results using different combinations (news content, social context,

TABLE 4.14: FakeNewsNet Dataset

News Source	News Articles	Fake News Articles	Number of Users
BuzzFeed	182	91	15257
PolitiFact	240	120	23865

and content+context) presented with the proposed approach. The recommended model obtained an accuracy of 86.84% and 89.19%, respectively, among the content and social context-based methods. Combining social-context and news-content features, the proposed model achieved a marginal improvement over the baseline methods with an accuracy of 92.30%. With these results, the effectiveness of social context-based features were recommended for fake news classification.

In this research, considering all classifiers' performance, a validation accuracy of 92.30% was achieved with the PolitiFact dataset with the proposed deep architecture. From figures 4.7 and 4.8, the classification results with the proposed deep neural network are shown, the validation accuracy was high, and cross-entropy loss was minimum using both real-world fake news dataset. The proposed model achieved accuracy with 91.80% using the BuzzFeed dataset (refer to figure 4.8). To validate the performance of the proposed model, more performance parameters included (False Positive Rate (FPR) and False Negative Rate (FNR)). The false-positive rate was 9.52%, and the false-negative rate was 13.64% with the proposed model using the BuzzFeed dataset (refer to table 4.24 for more details). The false-negative rate is 13.04%, and the false-negative rate was 9.52% with the proposed model using the PolitiFact dataset (refer to table 4.26 for more details). Results motivated the researchers to use the proposed method-EchoFakeD to classify fake news in their research.

4.2.4 Comparison with existing classification methods

From tables 4.27 and 4.28, a comparison between existing classification benchmarks with the proposed model (EchoFakeD) showed. Table 4.27 showed the classification results with BuzzFeed dataset, and table 4.28 showed the classification results with PolitiFact dataset. The FPR and FNR were also less with the proposed model. Existing studies primarily focused on the news content-based analysis. The problem of fake news has investigated the content-based

TABLE 4.15: Hyperparameters for EchoFakeD

Hyperparameter	**Value**
Number of dense layers	5
Number of hidden nodes	128,128,32,2
Activation function	ReLU
Loss function	Binary cross-entropy
Optimizer	Adam
Dropout	0.2
Learning rate	0.1
Number of epochs	20
Batch-size	64

TABLE 4.16: Dimensionality of Feature Matrices

Matrix	**Dimension**
News-user engagement matrix (U)	(182 x 15257)
Count matrix (N)	(182 x 1500)
User-community matrix (C)	(15257 x 81)
Tensor (T)	(182 x (15257 x 81)
Mode-1 tensor (X_1)	(182 x (15257 x 81))
Input Matrix (content + context)	(182 x 1503)

TABLE 4.17: Confusion matrix for news content-based classification with EchoFakeD (BuzzFeed)

	Predicted Positive	**Predicted Negative**
Actual Positive	17 (TP)	4 (FN)
Actual Negative	3 (FP)	16 (TN)

TABLE 4.18: Confusion matrix for social context-based classification with EchoFakeD (BuzzFeed)

	Predicted Positive	**Predicted Negative**
Actual Positive	18 (TP)	3 (FN)
Actual Negative	2 (FP)	16 (TN)

TABLE 4.19: Confusion matrix for news content + social context-based classification with EchoFakeD (BuzzFeed)

	Predicted Positive	**Predicted Negative**
Actual Positive	19 (TP)	3 (FN)
Actual Negative	2 (FP)	19 (TN)

TABLE 4.20: Confusion matrix for news content-based classification with EchoFakeD (PoitiFact)

	Predicted Positive	**Predicted Negative**
Actual Positive	17 (TP)	3 (FN)
Actual Negative	2 (FP)	16 (TN)

TABLE 4.21: Confusion matrix for social context-based classification with EchoFakeD (PoitiFact)

	Predicted Positive	**Predicted Negative**
Actual Positive	17 (TP)	2 (FN)
Actual Negative	2 (FP)	16 (TN)

TABLE 4.22: Confusion matrix using content and context-based features with EchoFakeD (PoitiFact)

	Predicted Positive	**Predicted Negative**
Actual Positive	19 (TP)	2 (FN)
Actual Negative	3 (FP)	20 (TN)

TABLE 4.23: Classification results with BuzzFeed

Approach	**Precision**	**Recall**	**F1-Score**	**Accuracy**
EchoFakeD with News Content	0.8500	0.8095	0.8293	0.8250
EchoFakeD with Social Context	0.8571	0.9000	0.8780	0.8718
EchoFakeD with Content+Context	**0.9047**	**0.8636**	**0.8837**	**0.9180**

TABLE 4.24: FPR and FNR using BuzzFeed

Approach	FPR	FNR
EchoFakeD with News Content	0.1579	0.1905
EchoFakeD with Social Context	0.1111	0.1429
EchoFakeD with Content+Context	**0.0952**	**0.1364**

TABLE 4.25: Classification results with PolitiFact

Approach	Precision	Recall	F1-Score	Accuracy
EchoFakeD with News Content	0.8500	0.8947	0.8718	0.8684
EchoFakeD with Social Context	0.8947	0.8947	0.8947	0.8919
EchoFakeD with Content+Context	**0.8636**	**0.9048**	**0.8837**	**0.9230**

TABLE 4.26: FPR and FNR using PolitiFact

Approach	FPR	FNR
EchoFakeD with News Content	0.1111	0.1500
EchoFakeD with Social Context	0.1111	0.1053
EchoFakeD with Content+Context	**0.1304**	**0.0952**

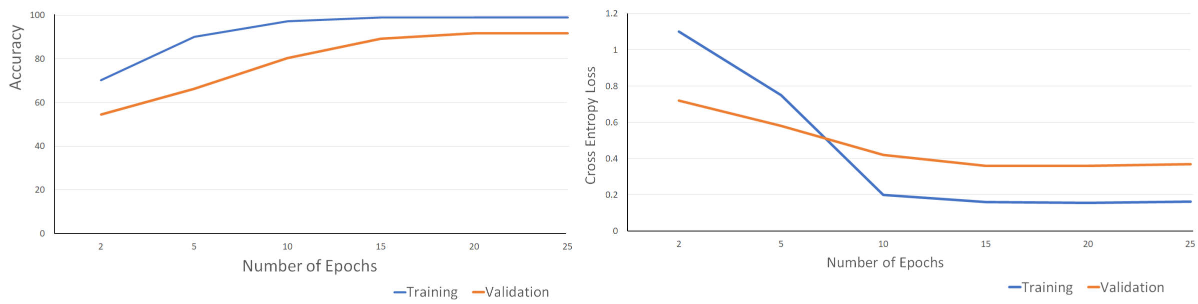

FIGURE 4.7: Classification accuracy and Cross-entropy Loss with EchoFakeD using BuzzFeed

TABLE 4.27: Comparison with existing benchmarks with BuzzFeed

Authors	Precision(%)	Recall(%)	F1-Score(%)
(Castillo et al. 2011)	73.50	78.30	75.60
(Castillo et al. 2011)-RST	79.50	78.40	78.90
(Gupta et al. 2018)-CITDetect	65.70	100.00	79.20
(Gupta et al. 2018)-CIMTDetect	72.90	92.30	81.30
(Papanastasiou et al. 2019)(CLASS-CP)	85.20	83.00	83.50
(Zhou & Zafarani 2018)	84.90	85.20	84.20
(Kaliyar et al. 2020a) (DNN-with echo chamber)	83.33	86.96	85.11
Proposed model-EchoFakeD	**90.47**	**86.36**	**88.37**

TABLE 4.28: Comparison with existing benchmarks using PolitiFact

Authors	Precision(%)	Recall(%)	F1-Score(%)
(Castillo et al. 2011)	77.70	79.10	78.30
(Castillo et al. 2011) -RST	82.30	79.20	79.30
(Gupta et al. 2018) -CITDetect	67.90	97.50	79.10
(Gupta et al. 2018)-CIMTDetect	80.30	84.20	81.80
(Papanastasiou et al. 2019)(CLASS-CP)	87.20	82.10	84.30
(Kaliyar et al. 2020a) (DNN-with echo chamber)	82.10	84.60	84.04
Proposed model-EchoFakeD	**86.36**	**90.48**	**88.37**

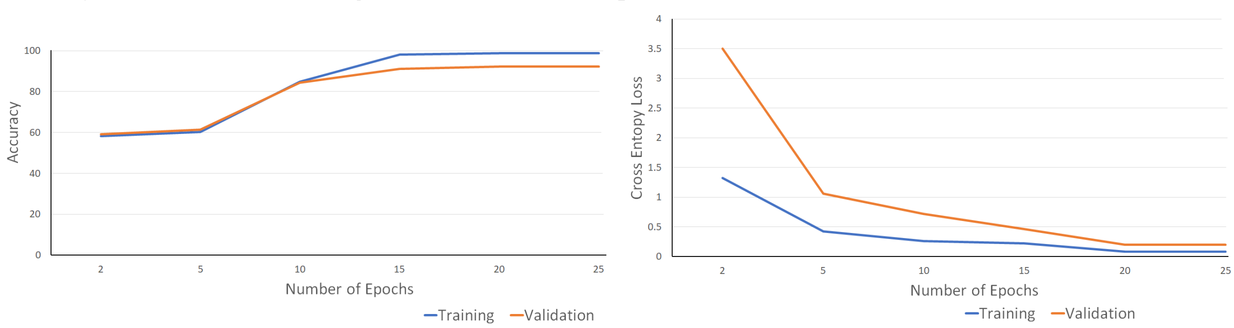

FIGURE 4.8: Classification Accuracy and Cross Entropy Loss with EchoFakeD using PolitiFact

attributes and the relationship between news articles and social media users. The proposed approach was one step ahead of the existing one. The proposed approach investigated the problem of fake news with an efficient deep neural network using the feature-vectors receiving from coupled matrix-tensor factorization method as a 3-mode tensor. In this approach, a tensor created using news articles' social context with several existing network communities. This method improved the performance of fake news classification compared to the existing methods. Results further motivated the researchers to use the proposed deep neural network compared to existing traditional methods for efficient results.

4.2.5 Discussion

In figure 4.9, an example available on social media of fake news showed. In this research, extensive feature set-based studies were performed for the classification of fake news. News content-based methods primarily focused on extracting different features from fake news articles, including both content-based (B) as well as style-based. Style-based strategies mainly concentrate on manipulators and creators' writing style (A) for the context of fake news. It was evident that for efficient fake news detection, content-based methodologies were alone, not sufficient. To investigate fake news articles with social context-based methods was the main necessity. Social context-based methods deal with the relationship among users, news article, and related publishers. These methodologies were efficient to recognize fake news articles. Social context (C) provides valuable information about users-based interaction with both fake news and real news. In the era of computing, at any social media platform, a user is always connected to a specific group of peoples having the same mindset or liking is called a user-community (D). These user communities can be an essential factor for fake news classification due to their common perception about sharing articles. Therefore, an effective Deep Neural Network combining (B+C+D) the content level features of news articles with user's social engagement (Echo-chamber infused) was designed to achieve significant results. Subsequently, the tensor factorization-based approach used with content as well as context-based information.

In this research, a user's engagement with the news articles captured and fused with user-community interaction to form a 3-mode tensor (content, social context, and user-community information). This tensor was capable of handling multi-relational data and provided a higher dimensional generalization of matrices. The tensor factorization based method decomposed the higher-order tensor into low-rank tensors (Khatri & Rao 1968). The resulting low-rank tensors capture the complicated relations between the different objects, represents as input tensors

FIGURE 4.9: An Example of Fake News (Source: Facebook)

for deep learning classification models. Therefore, in this research, a coupled matrix-tensor factorization method (Gupta et al. 2018) used. The proposed model (EchoFakeD) employed both contents and context-based information to validate the outcomes. The proposed model outperformed compared to the existing detection methods and obtained an accuracy of 92.30%.

4.3 Summary

A methodology for fake news detection that considers news content information collected from the text of news and social context information obtained from the echo chambers presented. The factorization method was evaluated with BuzzFeed as well as the PolitiFact dataset. A comparative analysis with news content, social context, and a combination of both were presented. It recognized that the combined content and context approach gave better results. A deep neural network further improved the classification results as compared to the XGBoost-a ensemble machine learning approach. For an effective classification, a tensor factorization approach presented with the proposed network. The performance of the model validated and verified with the dataset: BuzzFeed and PolitiFact. A detailed analysis showed different features: news content, social context, and the combination of news content and social context. With the classification results, it was clear that the combination approach using tensor factorization gave better performance, as shown in the evaluation parameters. The proposed system improved the classification result in terms of F1-score and accuracy compared to existing detection methods.

Chapter 5

A Hybrid Model for Effective Detection of Fake News with a Novel COVID-19 Dataset *

*The results presented in this chapter are published in: **Kaliyar, R.K.;** Goswami, A. and Narang, P. (2021). A Hybrid Model for Effective Fake News Detection with a Novel COVID-19 Dataset.In Proceedings of the 13th International Conference on Agents and Artificial Intelligence - Volume 2: ICAART, ISBN 978-989-758-484-8, pages 1066-1072. DOI: 10.5220/0010316010661072

5.1 Introduction

This investigation presented the research work with the proposed hybrid model for fake news identification using many convolutional layers with different kernel sizes following an LSTM network with three dense layers. A neural network with two convolutional layers was created having varying kernel sizes to learn the model with different word-sized vectors. The feature maps of CNN were constructed as valuable features which passed as the input of LSTM. Experimental results demonstrated the proposed hybrid model's effectiveness compared to other existing CNN and RNN-based classification models. The proposed model utilized the power of feature extraction using an advanced pre-trained word embedding model. The embedding layer produced the vectors for each word index and improved each word's embedding during training. The proposed model's novelty lies in designing a neural network with different sized kernels and filters in each convolutional layer. The convolutional layer's output was passed as an input to an LSTM layer. To make the C-LSTM model deeper, three dense layers were taken to enable the composition of features from lower layers, potentially modelling the data, quickly approaching the end goal and a higher-order decision boundary. The proposed model performed very well on PHEME and the FN-COV dataset with an accuracy of 91.88% and 98.62%, respectively. The proposed model obtained better classification results as compared to existing methods for fake news detection.

5.2 Experimental setup and methodology

In this section, various experiments and techniques presented that were utilized to achieve the research objectives along with the results of the experiments.

5.2.1 Dataset

In this subsection, the real-world fake news datasets used in this research discussed in details.

FN-COV

For creating the dataset, around 69,976 news articles were collected with 44.84% of fake in total from the GDELT project [†] supported by Google. It recently released with short parts of worldwide news coverage mentioning COVID-19 (Cui & Lee 2020). The collection included

[†] https://blog.gdeltproject.org

several topics that have been trending during COVID-19. For our experiment, we selected COVID19, quarantine, and social distancing tag related news articles. This dataset consists of five attributes: ‘Date’, ‘URL’, ’Title’, ‘Text’, and ‘Label’. The date corresponds to the published date of the news, ‘URL’ represents the web address of the published news, ‘Title’ represents the headline of the news, ‘Text’ represents the content of the news article, and ‘Label’ indicates whether the news is fake or not.

PHEME

PHEME dataset [‡] was a collection of tweets scraped from Twitter, posted at the time of breaking news having five events. The events included were: Charlie Hebdo shooting, from which we have 38,290 instances of news with 22% rumour content. Ferguson event dataset consists of 24,177 cases with 24.8% of rumour content. German wings plane crash comprises 4489 instances of tweet level text data in the set with 50.7% rumour. Ottawa shooting, which took place in Ottawa Parliament Hills in 2014, shall consist of 12,284 cases with 52.8% rumour and Sydney Siege where the gunmen took hostages at a cafe in Sydney in December 2014, consists of 24,001 instances tweet level text streams with 42.8% rumoured tweets.

5.2.2 Pre-processing

Text pre-processing is the practice of cleaning and preparing text data. In short, pre-processing relates to all the transformations on the unprocessed data before it supplied to the machine learning or deep learning algorithm. NLTK and re are standard Python libraries used to handle text pre-processing tasks. Such transformations are: Remove HTML tags, extra white spaces, special characters, convert to lowercase all input texts, convert several words to numeric form, and finally remove numbers.

5.2.3 Architecture of the proposed hybrid C-LSTM model

In Figure 5.1, the layered architecture of the proposed deep neural network showed. In the architecture, the first layer was an embedding layer that took the input as a vector of 1000 word indices of length 32 following by a convolutional layer that implements matrix multiplications-based operations (Collobert & Weston 2008; Sainath et al. 2015; Vasudevan et al. 2019; Yang et al. 2018). The first convolutional layer consists of kernel size=3 and 32 filters, followed by

[‡] http://www.zubiaga.org/datasets/

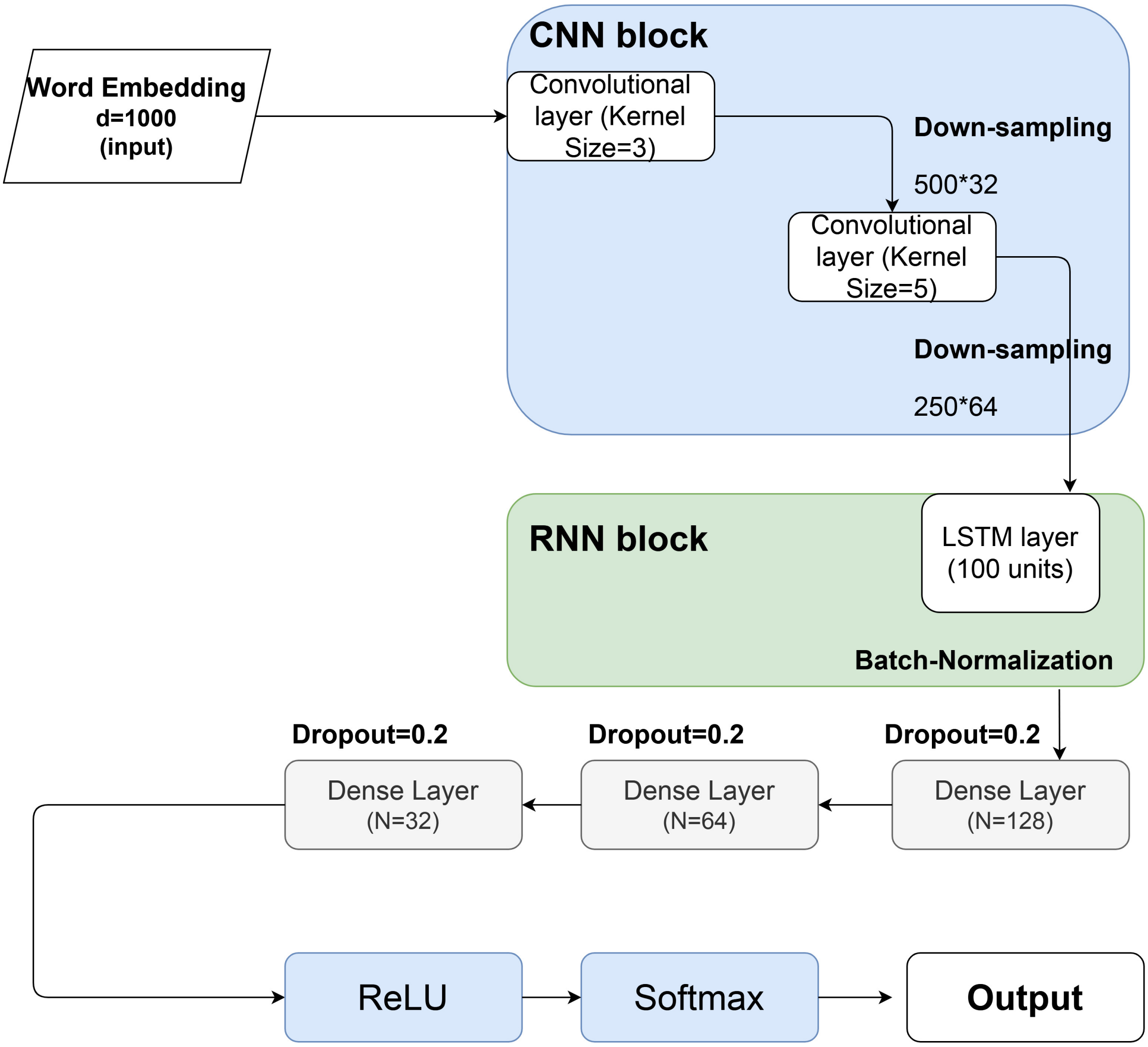

FIGURE 5.1: Proposed Model

TABLE 5.1: Optimal Hyper-parameters for the proposed hybrid model

Hyper-parameter	Description or Value
No. of Convolutional layer	2
No. of Max-pooling layer	2
No. of Kernel-sizes	3 and 5
No. of Dense layer	3
No. of filters in conv-layers	128,64,32
No. of filters in dense-layers	128,64,32,2
Loss function	binary cross-entropy
Activation function	ReLU
Optimizer	Adam
Metrics	Accuracy
Batch-size	128, 32
Batch-Normalization	Yes
Number of Epochs	20
Dropout	0.2

max-pooling. The second convolutional layer consists of kernel size=5 and 64 filters, followed by max-pooling. In the model, two pooling layer were taken (Vasudevan et al. 2019; Yang et al. 2018; Zhong et al. 2019) which effectively down-sampled the output of the last layer and reduced the number of operations required for all the following layers present in the network. The next layer in the architecture was an LSTM layer used to handle the nature of sequential data (Roy et al. 2018). This layer took complex word combinations as input and the length of several units as output. Next, a flatten layer was taken as a function that converted the features taken from the pooling layer and map them to a single column for further processing. Then, three dense layers were considered in the network architecture. The functionality of a dense layer as a linear operation (Vasudevan et al. 2019; Yang et al. 2018; Zhang et al. 2020) in which every input is connected to every output by some weight. The first dense layer has 128 nodes and a dropout of 0.2. With the small value of dropout, the accuracy will gradually increase, and loss will gradually decrease. We selected the dropout value because it helped reduce the classification model's complexity and prevent over-fitting (Vasudevan et al. 2019; Yang et al. 2018). The second dense layer also has 64 hidden nodes with a dropout of 0.2. The third dense layer has 32 hidden nodes and a dropout of 0.2. ReLU (Rectified Linear Unit) was taken as the activation function. It was capable enough to remove negative values from an activation map in a given network and increases the non-linear properties (Vasudevan et

TABLE 5.2: Comparison with Existing classification results using a real-world rumour dataset-PHEME

Author	Model	Accuracy (%)
(Zubiaga et al. 2016)	C-Random Forest	63.00%
(Zubiaga et al. 2016)	BOW+NB	68.15%
(Yu et al. 2017)	1-layer CNN	79.74%
(Zubiaga et al. 2016)	TF-IDF+KNN	80.94%
(Zubiaga et al. 2016)	BOW+DT	81.00%
(Ajao et al. 2018)	1-layer LSTM	82.76%
(Ajao et al. 2018)	LSTM-CNN	83.53%
(Ajao et al. 2018)	BILSTM-CNN	84.66%
(Ma et al. 2016, 2018)	RNN	86.12%
Our proposed model	**C-LSTM**	**91.88%**

TABLE 5.3: Performance of the C-LSTM model with PHEME and FN-COV

Dataset	Model	Precision	Recall	F1-Score
PHEME	C-LSTM	0.902	0.904	0.903
FN-COV	C-LSTM	0.992	0.989	0.994

al. 2019) of the decision-making function. The equation of ReLU as:

$$\sigma = max(0, z) \tag{5.1}$$

In this research, binary cross-entropy used as a loss function. For binary classification problems, the equation of cross-entropy can be defined as:

$$L = -(y log(p) + (1-y) log(1-p)) \tag{5.2}$$

If $M > 2$ (i.e. multi-class classification), a separate loss calculates for each class label per observation and sum the result:

$$-\sum_{c=1}^{M} y_{o,c} log\left(p_{o,c}\right) \tag{5.3}$$

Here, *log* - the natural *log*, y - binary indicator (0 or 1), c- class label, o-observation, p - predicted probability.

In the proposed neural network, Adam was considered as an optimizer. Optimal hyper-parameters were selected (see Table 5.1 for more details) for the proposed hybrid model. Hyperparameter optimization obtains a tuple of different hyperparameters that generates the best classification results with our model, minimizing a predefined loss function.

TABLE 5.4: Accuracy of the C-LSTM model with PHEME and FN-COV

Dataset	Model	Accuracy (%)
PHEME	C-LSTM	91.88
FN-COV	C-LSTM	98.62

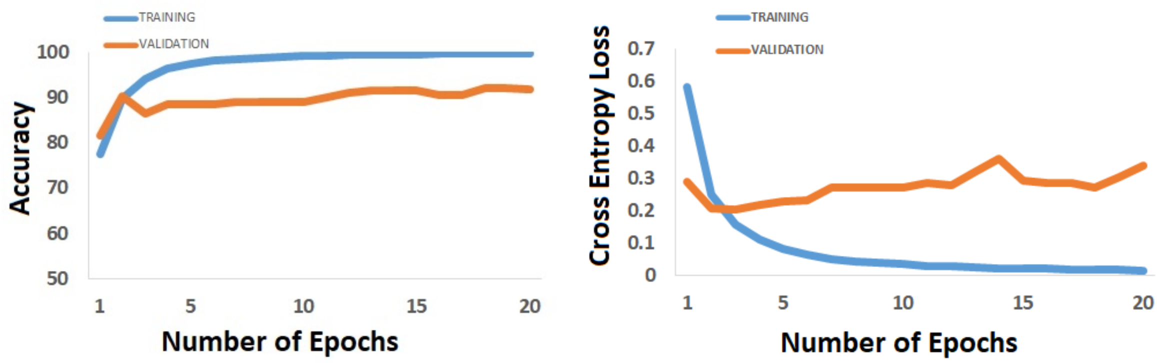

FIGURE 5.2: Accuracy and Cross Entropy Loss with C-LSTM using PHEME

In this research, the work was carried using the NVIDIA DGX-1 V100 machine. The machine was equipped with 40600 CUDA cores, 5120 tensor cores, 128 GB RAM and 1000 TFLOPS speed.

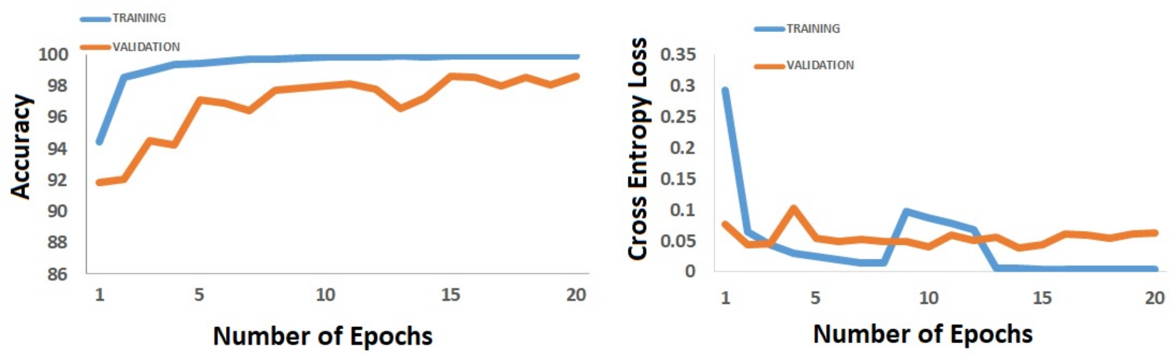

FIGURE 5.3: Accuracy and Cross Entropy Loss with C-LSTM using FN-COV

5.3 Results and discussion

Experimental and evaluation results tabulated in Table 5.2-5.4 using real-world fake news dataset: PHEME and the designed fake news dataset (FN-COV). The selection of optimal hyperparameters showed in Table 5.1. Classification results demonstrate that the recommended model performed more accurate results compared to other existing models for fake news detection.
From Table 5.4, shallow machine learning models have resulted in a maximum of 81% accuracy using the PHEME dataset. With deep learning models (CNN's, RNNs, etc.), the recurrent neural network architecture with GloVe pre-trained word embedding achieved 86.12% accuracy. The proposed models, which were deep and hybrid (a combination of both CNN and LSTM layers) in nature, performed exceptionally well and have resulted in more than 90% accuracy using the PHEME dataset. It also achieved an accuracy of 98.62% with the designed fake news dataset: FN-COV.
Figure 5.2 and 5.3 showed the accuracy and cross-entropy loss using PHEME and FN-COV dataset. It also traced the learning ability and generalizing power of our proposed model. The proposed model's performance over 20 epochs was quite remarkable on diverse and new dataset-FN-COV, respectively.
Cross-entropy loss was minimal in the case of FN-COV. In Table 5.2, various performance parameters were considered to validate the classification results. An F1-score with 99.40% with FN-COV and 90.30% with the PHEME dataset was achieved. Table 5.4, a comparison with existing classification results using a publicly available dataset (PHEME) showed. A 5% higher accuracy was achieved than the existing methods with the proposed hybrid model. The proposed model performed well with an accuracy of 98.62% using FN-COV. Motivated results were achieved with other real-world fake news datasets also.

5.4 Summary

In this research, with the proposed C-LSTM model, exemplary results were achieved. It captured both the temporal semantics and phrase-level representations and performed with optimized accuracy with minimal loss. Besides, a novel dataset was created of fake news propagating during COVID-19. The experimental results empirically showed the proposed model's effectiveness for fake news detection problem using PHEME and the FN-COV dataset.

Chapter 6

A Generalized Multichannel CNN for fake news detection *

*The results presented in this chapter are published in: **Kaliyar, R. K.**, Goswami, A., & Narang, P. (2021). MCNNet: Generalizing Fake News Detection with a Multichannel Convolutional Neural Network using a Novel COVID-19 Dataset. In 8th ACM IKDD CODS and 26th COMAD (pp. 437-437)

6.1 Introduction

In this chapter, the research work presented the classification results with the proposed multi-channel convolutional neural network to identify fake news effectively. Classification results showed that using different word embedding channels in a single network can provide valuable features for detecting fake news. The proposed model combined several parallel 1D-CNN's that read the source document using different kernel sizes. Kernel size represented the number of words considered in the convolution process across the whole input text document. The proposed model's learning ability processed at different n-grams at a moment. The proposed model determined how to combine these studies (different n-grams results) appropriately and how it affected model learning. Experimented were conducted with the proposed model using three real-world fake news datasets. The proposed model presented more accurate results using both existing and novel fake news dataset.

6.2 Experimental setup and methodology

In this section, various experiments and techniques presented that were utilized to achieve the research objectives along with the results of the experiments. The problem statement, dataset, methodology, and architecture of the proposed model also discussed in this section.

6.2.1 Dataset(s)

In this section, the datasets (refer to Table 6.1 for more details) used in this research discussed.

FN-COV

For creating the dataset, around 69,976 news articles were collected with 44.84% of fake in total from the GDELT project [†] supported by Google. The collection included several topics that had been trending during the initial phase of COVID-19. For the experiment, a selection was made of COVID19, quarantine, and social distancing tag related news articles. This dataset consists of five attributes: ‘Date’, ‘URL’, ’Title’, ‘Text’, and ‘Label’. The date corresponds to the published date of the news, ‘URL’ represents the web address of the published news, ‘Title’ represents the headline of the news, ‘Text’ represents the content of the information, and ‘Label’ indicates whether the news is fake or not.

[†] https://blog.gdeltproject.org

PHEME

PHEME [‡] dataset is a collection of tweets scraped from Twitter, posted at the time of breaking news having five events. The events included were: Charlie Hebdo shooting from which we have 38,290 instances of news with 22% rumour content. Ferguson event dataset consists of 24,177 instances with 24.8% of rumour content. German wings plane crash comprises 4489 examples of tweet level text data in the set with 50.7% rumour. Ottawa shooting, which took place in Ottawa Parliament Hills in 2014, shall consist of 12,284 samples with 52.8% rumour and Sydney Siege where the gunmen took hostages at a cafe in Sydney in December 2014 24,001 instances tweet level text streams with 42.8% rumour tweets.

CoAID

CoAID [§] is a real-world fake news dataset (Cui & Lee 2020) that recently introduced with the misinformation circulated during the pandemic of COVID-19, including fake news, along with users' social engagement. In this research, 2138 news were collected, 296,000 user engagements present in the news data, a total of 926 social media posts related to COVID-19, and ground truth output labels.

TABLE 6.1: Description of Dataset(s)

Dataset	Total Instances	Fake	Real
PHEME	103211	31384	38592
FN-COV	69976	31384	38592
CoAID	2138	549	1589

6.2.2 Pre-processing

Text pre-processing is the practice of cleaning and preparing text data. NLTK and re are standard Python libraries used to handle text pre-processing tasks. Such transformations are:

- Remove HTML tags, extra white spaces, special characters.
- Convert to lowercase all input texts.
- Convert several words into some numeric form.
- Remove numbers.

[‡] http://www.zubiaga.org/datasets/
[§] https://github.com/cuilimeng/CoAID

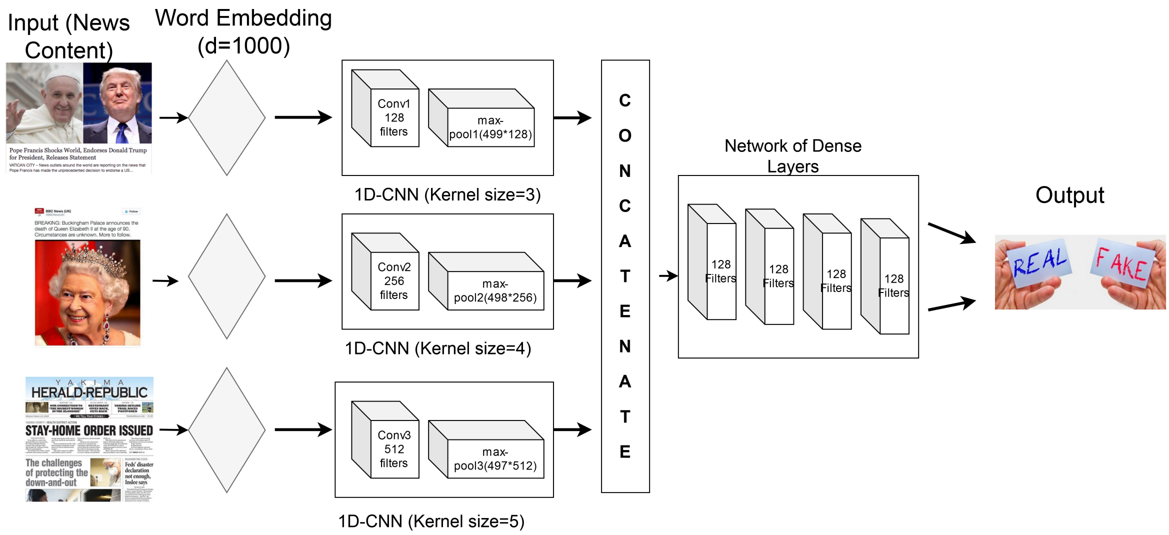

FIGURE 6.1: Proposed Model

6.2.3 Fake news detection methodologies

The great success of CNN's in computer vision motivated researchers to utilize the neural network architecture in different natural language processing tasks for extracting high-level n-gram features using different kernel sizes. Generally, First layer of 1D-CNN model transformed indices of a input sentence $s_1, s_2, \ldots, s_n$ tokens in *Vocab* into a series of vector $W_{s1}, W_{s2}, \ldots, W_{sn}$. $W_i \, \varepsilon \, R^d$ is the i^{th} column of embedding matrix W and d was the number of embedding space dimensions. The next layer in CNN's architecture was a convolutional layer in which each neuron (entry of z_{i+1}) received input from a square receptive field with $k \times k$ parameters. The whole layer only has k^2 parameters, responsible for convolving trainable filters (kernels) on the input data to extract valuable patterns. These kernels can create features based on the collocation of tokens in the input text. Next, the extracted features were fed into another layer, called the pooling layer, for down-sampling features by an operation like averaging or taking maximum. Finally, after these two layers, all extracted high-level features constructed a vector using a flattened layer.
This research combined three 1D-CNN into a single unified structure with variable kernel sizes and filters. The architecture of the proposed model was as:

The proposed Multichannel Convolutional Neural Network

The multi-channel CNN architecture consists of three parallel CNN's with different inputs called channels with varying types of information. Finally, all extracted feature vectors created a dense vector using the concatenated layer, and this vector fed into a fully connected network for making a prediction. Given the dimension of the reconstructed feature map $R''_{i,k}$ as $d \times k$, a neural network with kernel size $d \times h$ and the number of filters l is employed. Significantly, it generated one scalar feature as a product:

$$s_j = Relu(W_c.R''_{i,j:j+h-1} + b_c) \tag{6.1}$$

where w_c and b_c were the weights and the bias of the convolutional filter. In this approach, a feature vector $s_j \, \varepsilon \, R''^{l}$ is constructed with l filters. By replicating the convolution processes for each window of h, a sequence of standard feature vectors was achieved:

$$s = [s_1, s_2, \ldots, s_{k-h+1}] \tag{6.2}$$

Using a contextualized representation might not provide enough information about all textual features in the model to make an accurate decision. Thus, for developing a reliable fake news detection system, an approach was investigated to extract valuable features based on different news aspects by utilizing various types of text representation, including word embedding, kernel size, and filters in one architecture. Each of these representations illustrated distinct aspects of news articles. Figure 6.1 showed the overall architecture of the proposed model.

Consecutive three groups were conducted, convolutional layers, max-pooling layers, and word-document matrix to extract notable features based on the words that appear in news articles. All three-channel input was an advanced word embedding model like GloVe, which contains 100-dimensional vectors for 3 million word types. In the architecture, the first layer with the first channel was an embedding layer that accepted the input as a vector of 1000 word indices of length 32 following by a convolutional layer that performed matrix multiplications-based operations (Sainath et al. 2015; Vasudevan et al. 2019; Yang et al. 2018). The first layer consists of kernel size=3 and 32 filters, followed by max-pooling. In the second channel, the convolutional layer consisted of kernel size=4 and 64 filters, followed by max-pooling operations and so on. In each channel, one pooling was taken (Sainath et al. 2015; Vasudevan et al. 2019; Zhou & Zafarani 2019) which effectively down-sampled the output of the last layer and reduced the number of operations required for all the following layers present in the network.

Subsequently, three flatten layers were taken as a function that converted the features obtained from the pooling operation and map it to a single column that further transferred to the fully connected layer. The model was made with a network of four dense layers. The functionality of a dense layer as a linear operation (Roy et al. 2018; Sainath et al. 2015) in which every input was connected to every output by some weight. In the dropout, some nodes dropped randomly so that forward propagation and backward propagation will consider only retained nodes in that iteration. A random number generated for each node, and if it is less than $(1-p)$, where p was a probability parameter (say 0.2), then that node discarded. Else it retained. Thus forward pass equation with dropout at layer l will become:

$$\tilde{y}^{(l)} = Mask(p) * y^{(l)} \tag{6.3}$$

$$z_i^{(l+1)} = w_i^{(l+1)}.\tilde{y}^{(l)} + b_i^{(l+1)} \tag{6.4}$$

Here $Mask(p)$ was the mask for each node depending on a generated random number compared with the dropout rate. With a small value of dropout, the accuracy will gradually increase, and loss will gradually decrease. The value of dropout was selected because it helped to reduce the classification model's complexity and prevent over-fitting (Ghosh & Shah 2018; Shu et al. 2017). The embedding vectors encode a complete sentence processed by one of the leading models, which outputs a feature vector x that represented the whole sentence. This vector was then passed to a classification layer that applied $softmax$ function to estimate the predictive probabilities for all K labels:

$$p(y=k|X) = \frac{exp(w_k^T x + b_k)}{\sum_{k'=1}^{K} exp(w_{k'}^T x + b_{k'})} \tag{6.5}$$

The activation functions (z) applied element-wise; ReLU (Rectified Linear Unit) was taken as the activation function. It increased the non-linear properties of the decision-making function without affecting any other fields of the convolution layer. We can define the equation of ReLU as:

$$\sigma = max(0, z) \tag{6.6}$$

The loss functions are non-negative values that mainly used to estimate the variability between the predicted value $(\widehat{y})$ and the original value (y) and enhance the model's generalizing capability. The basic structure of the loss function is:

$$L(\theta) = \frac{1}{n}\sum_{i=1}^{n} L(y^{(i)}, f(x^{(i)}, \theta)) \tag{6.7}$$

where θ indicates the value of a parameter to the model, x indicates the model's feature matrix, and y denotes the model's actual labels. In this research, we employed categorical cross-entropy loss. Each prediction resembles the true class value in the cross-entropy loss function and a score calculated. The score further used to penalize the prediction's probability based on the real value difference. The equation for cross-entropy function was as below:

$$L(y,\widehat{y}) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{c}(y_{i,j}log(\widehat{y}_{i,j})) \tag{6.8}$$

where data samples ranging from 1 to N and the classes range from 1 to C. The term $y_{i,j}$ in the equation corresponds to the one-hot encoded label at the i^{th} index of the j^{th} category. And the term $\widehat{y}_{i,j}$ corresponds to the prediction of the model for the samples as i^{th} index. With the detection deadline k, the end goal of the optimization process was to find the optimal θ :

$$\hat{\theta} = arg\ min_{\theta}L(\theta,k) \tag{6.9}$$

In the proposed neural network, Adam was taken as an optimizer. Optimal hyper-parameters (which minimizes a predefined loss function) are shown with Table 6.2. This research work was carried with the NVIDIA DGX-1 V100 machine. The machine is furnished with 40600 CUDA cores, 5120 tensor cores, 128 GB RAM and 1000 TFLOPS speed.

TABLE 6.2: Optimal Hyper-parameters for the proposed Multichannel Neural Network

Hyper-parameter	Value
No. of Convolutional layer	3
No. of Max-pooling layer	3
No. of Kernel-sizes	3, 4, and 5
No. of Dense layer	4
No. of filters in conv-layers	128,64,32
No. of filters in dense-layers	128,128,128,128
Loss function	Categorical cross-entropy
Activation function	ReLU
Optimizer	Adam
Metrics	Accuracy
Number of Epochs	25
Dropout	0.2

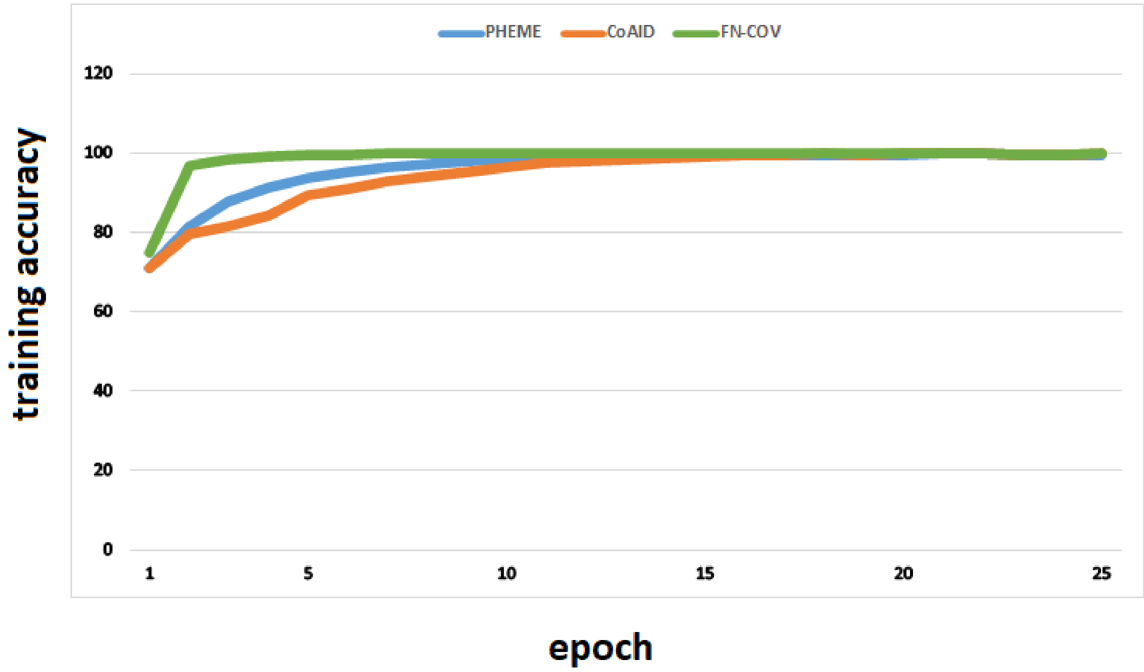

FIGURE 6.2: Accuracy with training samples

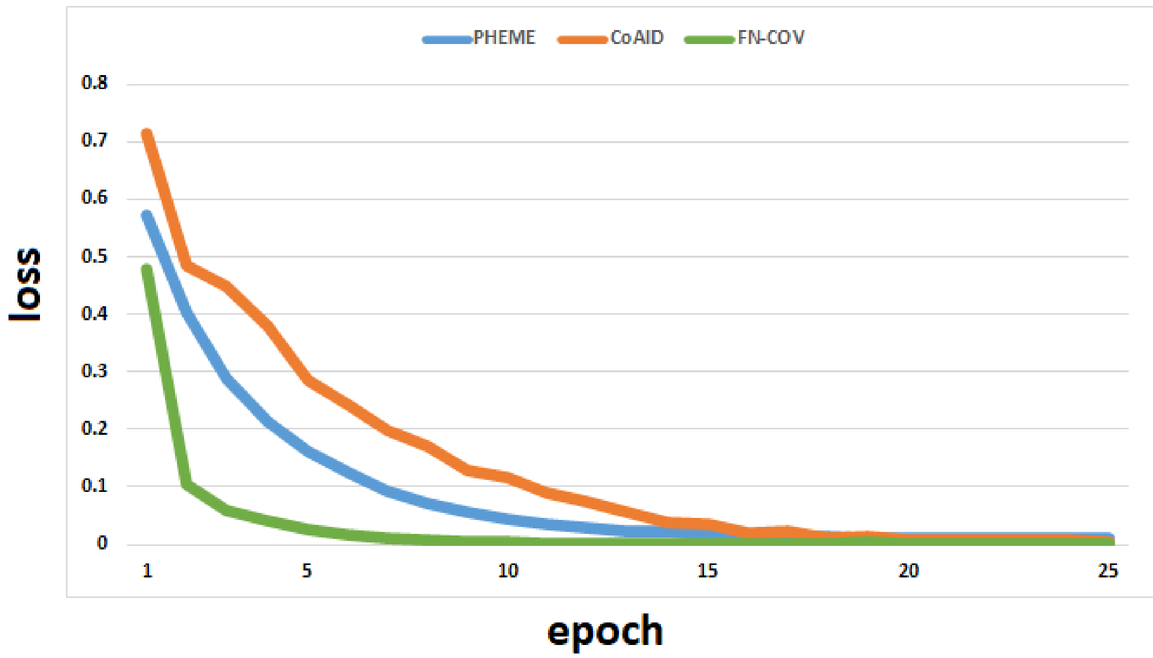

FIGURE 6.3: Cross Entropy loss with training samples

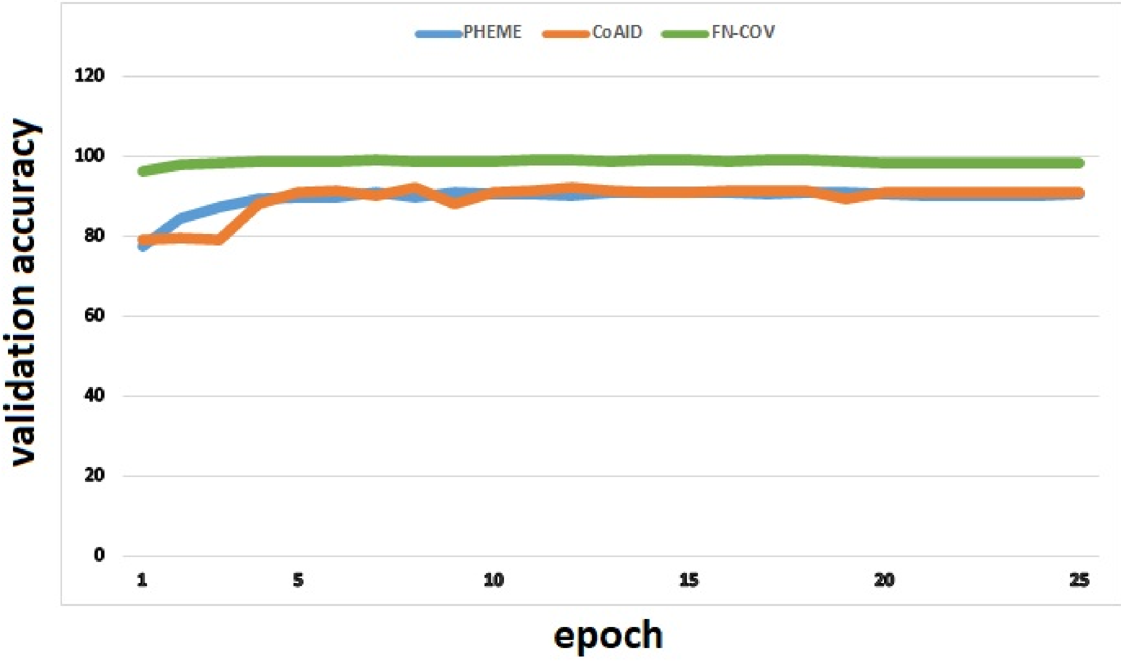

FIGURE 6.4: Accuracy with testing samples

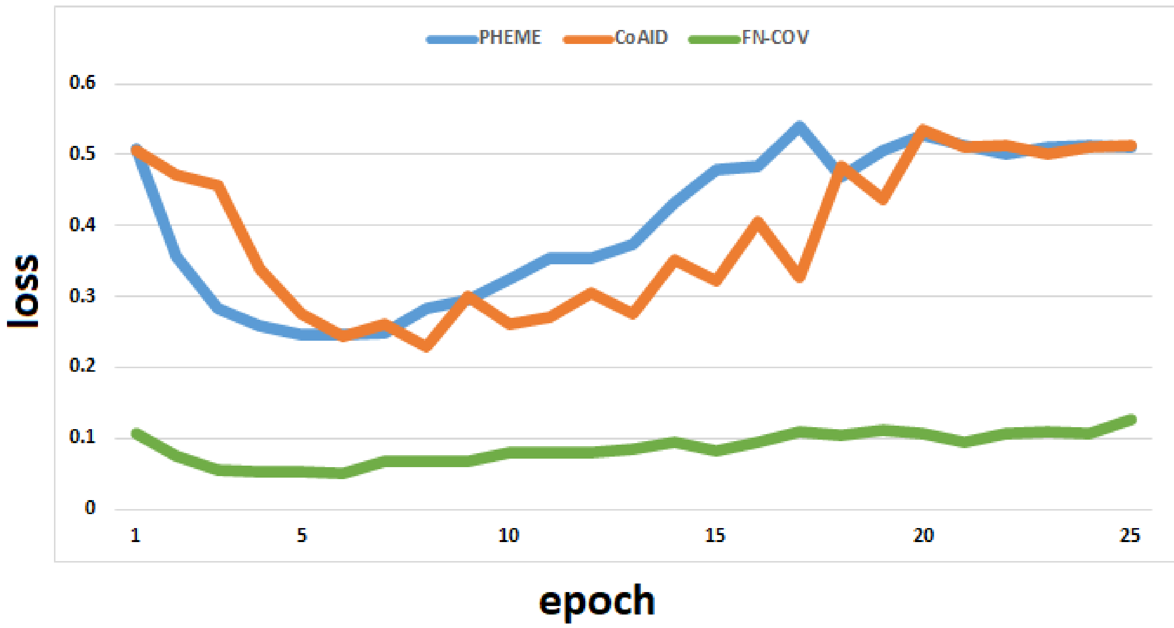

FIGURE 6.5: Cross Entropy loss with testing samples

TABLE 6.3: Classification Results

Dataset and Model	Precision	Recall	F1-Score	Accuracy
PHEME-AttCNN	0.826	0.810	0.818	0.827%
PHEME-MCNNet	0.866	0.830	0.848	0.903%
FN-COV-MCNNet	0.975	0.987	0.981	0.982%
CoAID-MCNNet	0.789	0.780	0.780	0.910%

TABLE 6.4: Comparison (the proposal model) with the existing detection methods using real-time rumour dataset-PHEME

Author	Model	Accuracy (%)
(Zubiaga et al. 2018a)	C-Random Forest	63.00%
(Zubiaga et al. 2018a)	BOW+NB	68.15%
(Yu et al. 2017)	1-layer CNN	79.74%
(Zubiaga et al. 2018a)	TF-IDF+KNN	80.94%
(Zubiaga et al. 2018a)	BOW+DT	81.00%
(Ajao et al. 2018)	1-layer LSTM	82.76%
(Ajao et al. 2018)	LSTM-CNN	83.53%
(Ajao et al. 2018)	BILSTM-CNN	84.66%
(Ma et al. 2016)	RNN	86.12%
The proposed model	**Multichannel CNN**	**90.30%**

6.3 Results and evaluations

Experimental and evaluation results tabulated in Table 6.3 and 6.4 using real-world fake news dataset: PHEME, CoAID and the designed fake news dataset (FN-COV). The selection of optimal hyperparameters showed in Table 6.2. Classification results were demonstrated the performance of the proposed model compared to other existing detection models for fake news.
Table 6.4 showed the classification results with shallow machine learning models produced a maximum of 81% accuracy using the PHEME dataset. With deep learning models (CNN's, RNNs, etc.), the recurrent neural network architecture with glove pre-trained word embedding achieved 86.12% accuracy. The proposed models, which was a combination of three 1D-CNN's are deep, have performed exceptionally well and resulted in more than 90% accuracy using the PHEME dataset. It also achieved an accuracy of 98.62% with the designed fake news dataset: FN-COV and 91.00% with CoAID. Figure 6.2 and 6.3 showed the accuracy and cross-entropy loss using PHEME, CoAID and FN-COV datasets. It also traced the learning ability and generalizing power of the proposed model. The proposed model's performance over 25 epochs was quite remarkable on diverse and new dataset-FN-COV, respectively.

Cross-entropy loss was minimal in the case of FN-COV. In Table 6.3, numerous performance parameters were considered to validate the model's performance. The F1-score with 99.40% with FN-COV and 90.30% with the PHEME dataset was achieved. Table 6.4, comparing existing classification results using a publicly available dataset (PHEME)showed. A 5% higher accuracy was obtained by the proposed approach compared to the existing systems. The proposed model performed well with an accuracy of 98.20% using FN-COV. Classification results demonstrated the proposed model's generalizing power using different performance parameters (Precision, Recall, F1-Score, Accuracy, etc.). Encouraging results were achieved with the proposed model (MCNNet), with more than 90% accuracy on all datasets, the best being 98.2% with FN-COV.

6.4 Summary

This research presented a multichannel convolutional neural network for effectively detecting fake news distributed online through web-based outlets. A novel dataset was created that contains fake news circulating widely during the pandemic of covid-19. The proposed model's performance was validated with different real-world fake news dataset: FN-COV, CoAID, and PHEME. The classification performance of the model was lucrative towards any fake news dataset. Ultimately, a significant improvement was achieved in detecting false information spread in social media.

Chapter 7

Conclusion and Scope of future work

7.1 Salient features and key findings

The goal of the research was to utilize the power of deep leaning for improving fake news detection. Hence, in this section, critical findings and salient features discussed that were used in the implemented models.

7.1.1 FNDNet and BERT-based deep learning approach

In the investigation, a look around exhibited that the presented approach did not rely on extracting hand-crafted features (Kumar & Shah 2018). Instead, the approach (FNDNet) outlined to learn the discriminatory features (Zhong et al. 2019) present in news articles automatically. The proposed model showed outstanding performance on large-scale real-world fake news datasets compared to existing detection approaches. The proposed model also focused on selecting an optimal depth of CNN's for collectivity accurate text classification problem. The classification results outperformed the existing implementations for fake news detection. Using the proposed approach, improved outcomes were obtained as compared to the baseline approaches which made FNDNet a promising model for accurate identification of fake news.
In another part of our research, a BERT-based in-depth learning approach (FakeBERT) was presented by fusing parallel blocks of the single-layer CNN's with BERT (Bidirectional Encoder Representations from Transformers). BERT was employed as a sentence encoder, which can get the context representation of a sentence. This research work distinguishes with previous research works (De Sarkar et al. 2018) where researchers examined a text sequence in a unidirectional way. Most of the existing detection methods (De Sarkar et al. 2018; Malik et al. 1991) presented with sequential neural networks to encode the relevant information. However, a deep neural network with a bidirectional training approach was an optimal and reliable solution for detecting fake news effectively. The proposed method enhanced real-time fake news detection performance with the semantic and long-distance dependencies present in text-based news articles.

7.1.2 Tensor decomposition-based deep neural networks (DeepFakE and EchoFakeD)

This research presented an effective deep learning model with both news content and context-related features. In this research, for textual modality and effective detection, an extensive feature set-based studies were performed to classify fake news. In the exploration, the user's engagement with the news articles was captured and fused with user-community interaction to form a 3-mode tensor (content, social context, and user-community information). This tensor handled multi-relational data (Hosseinimotlagh & Papalexakis 2018; Rabanser et al. 2017) and provided a higher dimensional generalization of input matrices. The tensor factorization method decomposed the higher-order tensor (Gupta et al. 2012; Gupta et al. 2018) into low-rank tensors. The resulting low-rank tensors capture the complex relations between the objects representing a graph-like structure in the dataset. The dimension of a combined matrix (content-context information) is 182x1503 was achieved, in which many news stories were 182, and the size of the input word embedding was 1503. With a coupled matrix-tensor factorization method (also known as CP-decomposition), the standard factorization method was utilized to decompose the user-level input matrix. In labelled data, the class information could help the factorization process identify fake news better. The architecture performed excellently on small and large real-world fake news datasets and reduced classification error. Using user community-based features with news content as a sizable dimensional tensor, the optimal results were achieved with a neural system having five hidden layers.

7.1.3 A Hybrid model for fake news detection

This research presented a hybrid model using convolutional layers with different kernel sizes with LSTM layers followed by three dense layers. A neural network was designed using two convolutional layers with varying kernel sizes to learn the model with different word-sized vectors. In our designed CNN, the feature maps constructed as a sequential feature that passed as an input of the developed model. This architecture organized each sentence into successive input features to help unravel variations within the same sentences. Experimental results demonstrated the proposed hybrid model's effectiveness compared to other existing CNN and RNN networks. The proposed model utilized the power of feature extraction using an advanced pre-trained word embedding model. In terms of the model's novelty, we used different kernel sizes and filters in each convolutional layer that passes input to the LSTM model. To make the C-LSTM model deeper and effective, three dense layers were took to enable the composition of features from lower layers, potentially modelling the data, approaching the end goal quickly, and a higher-order decision boundary. The proposed model performed very well on PHEME and the FN-COV dataset with an accuracy of 91.88% and 98.62%, respectively. The designed model achieved more accurate results as compared to existing approaches for the detection of fake news.

7.1.4 A generalized model for fake news detection

In this research, a multichannel convolutional neural network was presented to detect fake news effectively. It was demonstrated that using different word embedding channels with distinct sizes kernel in convolutional layers provided valuable features for detecting fake news. the proposed model combined multiple parallel CNN's that studied the source document employing different kernel sizes. The model's learning ability processed at different n-grams (groups of words in input word sentence) at a time. The model determined how to integrate these interpretations (different n-grams) best and how it affected model learning. Experimented were carried to validate the achievement of the designed system with three real-world fake news datasets. The proposed model showed accurate results using both existing and novel fake news dataset.

7.2 Scope of future work

After a detailed examination with the machine and deep learning models, the following topics/suggestions/methods identified as potential and promising directions for future research:

1. **Design an effective deep CNN for text and images in a combination**: Build a model for more effective and accurate detection of fake news in a real-time scenario (news content + images + metadata). Such detection methods can help future researchers incorporate multiple metadata information for more accurate news articles. A graph-based analysis to find fake news articles' exact propagation path of a news article's after some time duration can be explored using the above mentioned deep convolutional model.

2. **Design a deep learning model with bi-directional word embedding for news content and images both together**: Building a deep learning model (for news content + images) to detect fake news articles using multiple domains dataset. For future research work, such a detection method can help to investigate metadata information about the users and their connections in the news-user graph. The temporal features of news articles for tracking the propagation path and social media posts' statistics to construct a consolidated feature vector can detect fake news more effectively when deals with multiple output class dataset.

3. **Develop an effective detection model combining numerous features (Graph-based + user-level features + different communities + temporal characteristics)**: Develop a valuable deep learning model for effective fake news detection to examine the propagation

of news articles. For future research, a detailed interpretation can help solve fake news from different echo chambers present in social media data, which can be considered a group of personalities having the same opinion for any social concern. The prime motive to examine echo chambers in fake news detection implies that every user is co-related in a graph like structure on any social media platform like a community. It can help to examine new techniques of expressing the available information with tensors at a methodological level. As further research, a detection method can examine how the proposed model (EchoFakeD) improved the classification results while more features added.

4. **Develop a hybrid model for tracking the instances of news articles with both the binary and multi-class real-world fake news datasets:** Incorporating temporal level features from news articles available on social media with content and context level information and implementing for both the binary and multi-class real-world fake news datasets for enhanced and accurate classification. This type of hybrid approach can be valuable to identify fake news instances during the proliferation of a news article in a connected graph using multi-label output datasets.

5. **A generalized model for fake news detection toward multiple output label-based real-world dataset**: To design a generalized model which can provide a more accurate classification of fake news articles using multiple output label real-world fake news datasets. It can also be helpful with a different domain and language-based real-world fake news datasets for effective results. Users' credibility analysis can also be a potential research task with a generalized model for future researchers.

References

Abedalla, Ayat, Al-Sadi, Aisha, & Abdullah, Malak (2019). “A Closer Look at Fake News Detection: A Deep Learning Perspective”. *Proceedings of the 2019 3rd International Conference on Advances in Artificial Intelligence*, pp. 24–28.

Acar, Evrim, Kolda, Tamara G, & Dunlavy, Daniel M (2011). “All-at-once optimization for coupled matrix and tensor factorizations”. *arXiv preprint arXiv:1105.3422*.

Afroz, Sadia, Brennan, Michael, & Greenstadt, Rachel (2012). “Detecting hoaxes, frauds, and deception in writing style online”. *2012 IEEE Symposium on Security and Privacy*. IEEE, pp. 461–475.

Agarwal, Arush & Dixit, Akhil (2020). “Fake News Detection: An Ensemble Learning Approach”. *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE, pp. 1178–1183.

Aggarwal, Charu C (2014). *Data classification: algorithms and applications*. CRC press.

Ahmed, Hadeer, Traore, Issa, & Saad, Sherif (2017). “Detection of online fake news using N-gram analysis and machine learning techniques”. *International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*. Springer, pp. 127–138.

Ahmed, Hadeer, Traore, Issa, & Saad, Sherif (2018). “Detecting opinion spams and fake news using text classification”. *Security and Privacy* 1.1, e9.

Ajao, Oluwaseun, Bhowmik, Deepayan, & Zargari, Shahrzad (2018). “Fake news identification on twitter with hybrid cnn and rnn models”. *Proceedings of the 9th international conference on social media and society*, pp. 226–230.

Al-Barazanchi, Hussein A, Qassim, Hussam, & Verma, Abhishek (2016). “Novel CNN architecture with residual learning and deep supervision for large-scale scene image categorization”.

2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). IEEE, pp. 1–7.

Allcott, Hunt & Gentzkow, Matthew (2017). "Social media and fake news in the 2016 election". *Journal of economic perspectives* 31.2, pp. 211–36.

Anoop, K, Gangan, Manjary P, Deepak, P, & Lajish, VL (2019). "Leveraging heterogeneous data for fake news detection". *Linking and Mining Heterogeneous and Multi-view Data*. Springer, pp. 229–264.

Antoniadis, Sotirios, Litou, Iouliana, & Kalogeraki, Vana (2015). "A model for identifying misinformation in online social networks". *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*. Springer, pp. 473–482.

Araque, Oscar, Corcuera-Platas, Ignacio, Sanchez-Rada, J Fernando, & Iglesias, Carlos A (2017). "Enhancing deep learning sentiment analysis with ensemble techniques in social applications". *Expert Systems with Applications* 77, pp. 236–246.

Asparouhov, Tihomir & Muthén, Bengt (2010). "Weighted least squares estimation with missing data". *Mplus Technical Appendix* 2010, pp. 1–10.

Bai, Na, Wang, Zhixiao, & Meng, Fanrong (2020). "A Stochastic Attention CNN Model for Rumor Stance Classification". *IEEE Access* 8, pp. 80771–80778.

Basak, Rajesh, Sural, Shamik, Ganguly, Niloy, & Ghosh, Soumya K (2019). "Online Public Shaming on Twitter: Detection, Analysis, and Mitigation". *IEEE Transactions on Computational Social Systems* 6.2, pp. 208–220.

Batrinca, Bogdan & Treleaven, Philip C (2015). "Social media analytics: a survey of techniques, tools and platforms". *Ai & Society* 30.1, pp. 89–116.

Beltagy, Iz, Cohan, Arman, & Lo, Kyle (2019). "Scibert: Pretrained contextualized embeddings for scientific text". *arXiv preprint arXiv:1903.10676*.

Berant, Jonathan & Liang, Percy (2014). “Semantic parsing via paraphrasing”. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1415–1425.

Bhatt, Gaurav, Sharma, Aman, Sharma, Shivam, Nagpal, Ankush, Raman, Balasubramanian, & Mittal, Ankush (2017). “On the benefit of combining neural, statistical and external features for fake news identification”. *arXiv preprint arXiv:1712.03935*.

Bhatt, Gaurav, Sharma, Aman, Sharma, Shivam, Nagpal, Ankush, Raman, Balasubramanian, & Mittal, Ankush (2018). “Combining Neural, Statistical and External Features for Fake News Stance Identification”. *Companion of the The Web Conference 2018 on The Web Conference 2018*. International World Wide Web Conferences Steering Committee, pp. 1353–1357.

Blondel, Vincent D, Guillaume, Jean-Loup, Lambiotte, Renaud, & Lefebvre, Etienne (2008). “Fast unfolding of communities in large networks”. *Journal of statistical mechanics: theory and experiment* 2008.10, P10008.

Bondielli, Alessandro & Marcelloni, Francesco (2019). “A survey on fake news and rumour detection techniques”. *Information Sciences* 497, pp. 38–55.

Bozarth, Lia, Saraf, Aparajita, & Budak, Ceren (2020). “Higher ground? How groundtruth labeling impacts our understanding of fake news about the 2016 US presidential nominees”. *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 14, pp. 48–59.

Brummette, John, DiStaso, Marcia, Vafeiadis, Michail, & Messner, Marcus (2018). “Read All About It: The Politicization of “Fake News” on Twitter”. *Journalism & Mass Communication Quarterly* 95.2, pp. 497–517.

Cai, Chiyu, Li, Linjing, & Zeng, Daniel (2017). “Detecting social bots by jointly modeling deep behavior and content information”. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 1995–1998.

Cai, Deng, He, Xiaofei, Wu, Xiaoyun, & Han, Jiawei (2008). “Non-negative matrix factorization on manifold”. *2008 Eighth IEEE International Conference on Data Mining*. IEEE, pp. 63–72.

Caliskan, Aylin, Bryson, Joanna J, & Narayanan, Arvind (2017). "Semantics derived automatically from language corpora contain human-like biases". *Science* 356.6334, pp. 183–186.

Camacho-Collados, José, Pilehvar, Mohammad Taher, & Navigli, Roberto (2016). "Nasari: Integrating explicit knowledge and corpus statistics for a multilingual representation of concepts and entities". *Artificial Intelligence* 240, pp. 36–64.

Castillo, Carlos, Mendoza, Marcelo, & Poblete, Barbara (2011). "Information credibility on twitter". *Proceedings of the 20th international conference on World wide web*. ACM, pp. 675–684.

Cerisara, Christophe, Kral, Pavel, & Lenc, Ladislav (2018). "On the effects of using word2vec representations in neural networks for dialogue act recognition". *Computer Speech & Language* 47, pp. 175–193.

Chen, Guoguo, Parada, Carolina, & Heigold, Georg (2014). "Small-footprint keyword spotting using deep neural networks". *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, pp. 4087–4091.

Chen, James F & Wang, Jieh-Shan (1998). *Token distribution, registration, and dynamic configuration of user entitlement for an application level security system and method*. US Patent 5,784,463.

Chen, Tianqi & Guestrin, Carlos (2016). "Xgboost: A scalable tree boosting system". *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. ACM, pp. 785–794.

Chen, Weiling, Zhang, Yan, Yeo, Chai Kiat, Lau, Chiew Tong, & Lee, Bu Sung (2018). "Unsupervised rumor detection based on users' behaviors using neural networks". *Pattern Recognition Letters* 105, pp. 226–233.

Chen, Yimin, Conroy, Niall J, & Rubin, Victoria L (2015). "Misleading online content: Recognizing clickbait as false news". *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection*. ACM, pp. 15–19.

Cho, Kyunghyun, Van Merriënboer, Bart, Gulcehre, Caglar, Bahdanau, Dzmitry, Bougares, Fethi, Schwenk, Holger, & Bengio, Yoshua (2014). "Learning phrase representations using RNN encoder-decoder for statistical machine translation". *arXiv preprint arXiv:1406.1078*.

Chong, Eunsuk, Han, Chulwoo, & Park, Frank C (2017). "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies". *Expert Systems with Applications* 83, pp. 187–205.

Chu, Zi, Gianvecchio, Steven, Wang, Haining, & Jajodia, Sushil (2012). "Detecting automation of twitter accounts: Are you a human, bot, or cyborg?" *IEEE Transactions on dependable and secure computing* 9.6, pp. 811–824.

Clauset, Aaron, Newman, Mark EJ, & Moore, Cristopher (2004). "Finding community structure in very large networks". *Physical review E* 70.6, p. 066111.

Collobert, Ronan & Weston, Jason (2008). "A unified architecture for natural language processing: Deep neural networks with multitask learning". *Proceedings of the 25th international conference on Machine learning*. ACM, pp. 160–167.

Collobert, Ronan, Weston, Jason, Bottou, Léon, Karlen, Michael, Kavukcuoglu, Koray, & Kuksa, Pavel (2011). "Natural language processing (almost) from scratch". *Journal of machine learning research* 12.ARTICLE, pp. 2493–2537.

Conroy, Niall J, Rubin, Victoria L, & Chen, Yimin (2015). "Automatic deception detection: Methods for finding fake news". *Proceedings of the Association for Information Science and Technology* 52.1, pp. 1–4.

Corina, David P, Loudermilk, Brandon C, Detwiler, Landon, Martin, Richard F, Brinkley, James F, & Ojemann, George (2010). "Analysis of naming errors during cortical stimulation mapping: implications for models of language representation". *Brain and language* 115.2, pp. 101–112.

Cui, Limeng & Lee, Dongwon (2020). "CoAID: COVID-19 Healthcare Misinformation Dataset". *arXiv preprint arXiv:2006.00885*.

Dahl, George E, Yu, Dong, Deng, Li, & Acero, Alex (2011). "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition". *IEEE Transactions on audio, speech, and language processing* 20.1, pp. 30–42.

De Sarkar, Sohan, Yang, Fan, & Mukherjee, Arjun (2018). "Attending sentences to detect satirical fake news". *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 3371–3380.

Della Vedova, Marco L, Tacchini, Eugenio, Moret, Stefano, Ballarin, Gabriele, DiPierro, Massimo, & Alfaro, Luca de (2018). "Automatic online fake news detection combining content and social signals". *2018 22nd conference of open innovations association (FRUCT)*. IEEE, pp. 272–279.

Denil, Misha, Demiraj, Alban, Kalchbrenner, Nal, Blunsom, Phil, & Freitas, Nando de (2014). "Modelling, visualising and summarising documents with a single convolutional neural network". *arXiv preprint arXiv:1406.3830*.

Devlin, Jacob, Chang, Ming-Wei, Lee, Kenton, & Toutanova, Kristina (2018). "Bert: Pre-training of deep bidirectional transformers for language understanding". *arXiv preprint arXiv:1810.04805*.

Dharawat, Arkin, Lourentzou, Ismini, Morales, Alex, & Zhai, ChengXiang (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*.

Djidjev, Hristo N (2006). "A scalable multilevel algorithm for graph clustering and community structure detection". *International workshop on algorithms and models for the web-graph*. Springer, pp. 117–128.

Dong, Manqing, Yao, Lina, Wang, Xianzhi, Benatallah, Boualem, Sheng, Quan Z, & Huang, Hao (2018). "Dual: A deep unified attention model with latent relation representations for fake news detection". *International Conference on Web Information Systems Engineering*. Springer, pp. 199–209.

Dougherty, James, Kohavi, Ron, & Sahami, Mehran (1995). "Supervised and unsupervised discretization of continuous features". *Machine Learning Proceedings 1995*. Elsevier, pp. 194–202.

Egele, Manuel, Stringhini, Gianluca, Kruegel, Christopher, & Vigna, Giovanni (2013). "Compa: Detecting compromised accounts on social networks." *NDSS*.

Egele, Manuel, Stringhini, Gianluca, Kruegel, Christopher, & Vigna, Giovanni (2015). "Towards detecting compromised accounts on social networks". *IEEE Transactions on Dependable and Secure Computing* 14.4, pp. 447–460.

Farajtabar, Mehrdad, Yang, Jiachen, Ye, Xiaojing, Xu, Huan, Trivedi, Rakshit, Khalil, Elias, Li, Shuang, Song, Le, & Zha, Hongyuan (2017). "Fake news mitigation via point process based intervention". *International Conference on Machine Learning*. PMLR, pp. 1097–1106.

Faustini, Pedro Henrique Arruda & Covões, Thiago Ferreira (2020). "Fake news detection in multiple platforms and languages". *Expert Systems with Applications* 158, p. 113503.

Fazil, Mohd & Abulaish, Muhammad (2018). "A hybrid approach for detecting automated spammers in twitter". *IEEE Transactions on Information Forensics and Security* 13.11, pp. 2707–2719.

Feng, Song, Banerjee, Ritwik, & Choi, Yejin (2012). "Syntactic stylometry for deception detection". *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*. Association for Computational Linguistics, pp. 171–175.

Fernandez, Miriam & Alani, Harith (2018). "Online Misinformation: Challenges and Future Directions". *Companion of the The Web Conference 2018 on The Web Conference 2018*. International World Wide Web Conferences Steering Committee, pp. 595–602.

Ferrara, Emilio, Varol, Onur, Davis, Clayton, Menczer, Filippo, & Flammini, Alessandro (2016). "The rise of social bots". *Communications of the ACM* 59.7, pp. 96–104.

Founta, Antigoni Maria, Chatzakou, Despoina, Kourtellis, Nicolas, Blackburn, Jeremy, Vakali, Athena, & Leontiadis, Ilias (2019). "A unified deep learning architecture for abuse detection". *Proceedings of the 10th ACM Conference on Web Science*. ACM, pp. 105–114.

Fourney, Adam, Racz, Miklos Z, Ranade, Gireeja, Mobius, Markus, & Horvitz, Eric (2017). "Geographic and Temporal Trends in Fake News Consumption During the 2016 US Presidential Election." *CIKM*. Vol. 17, pp. 6–10.

Fu, Xianghua, Liu, Wangwang, Xu, Yingying, & Cui, Laizhong (2017). "Combine HowNet lexicon to train phrase recursive autoencoder for sentence-level sentiment analysis". *Neurocomputing* 241, pp. 18–27.

Ghanem, Bilal, Rosso, Paolo, & Rangel, Francisco (2018). "Stance Detection in Fake News A Combined Feature Representation". *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, pp. 66–71.

Ghani, Norjihan Abdul, Hamid, Suraya, Hashem, Ibrahim Abaker Targio, & Ahmed, Ejaz (2019). "Social media big data analytics: A survey". *Computers in Human Behavior* 101, pp. 417–428.

Ghosh, Souvick & Shah, Chirag (2018). "Towards automatic fake news classification". *Proceedings of the Association for Information Science and Technology* 55.1, pp. 805–807.

Giachanou, Anastasia, Ríssola, Esteban A, Ghanem, Bilal, Crestani, Fabio, & Rosso, Paolo (2020). "The role of personality and linguistic patterns in discriminating between fake news spreaders and fact checkers". *International Conference on Applications of Natural Language to Information Systems*. Springer, pp. 181–192.

Giasemidis, Georgios, Singleton, Colin, Agrafiotis, Ioannis, Nurse, Jason RC, Pilgrim, Alan, Willis, Chris, & Greetham, Danica Vukadinovic (2016). "Determining the veracity of rumours on Twitter". *International Conference on Social Informatics*. Springer, pp. 185–205.

Giatsoglou, Maria, Vozalis, Manolis G, Diamantaras, Konstantinos, Vakali, Athena, Sarigiannidis, George, & Chatzisavvas, Konstantinos Ch (2017). "Sentiment analysis leveraging emotions and word embeddings". *Expert Systems with Applications* 69, pp. 214–224.

Goldani, Mohammad Hadi, Momtazi, Saeedeh, & Safabakhsh, Reza (2021). "Detecting fake news with capsule neural networks". *Applied Soft Computing* 101, p. 106991.

Gorrell, Genevieve, Bontcheva, Kalina, Derczynski, Leon, Kochkina, Elena, Liakata, Maria, & Zubiaga, Arkaitz (2018). "Rumoureval 2019: Determining rumour veracity and support for rumours". *arXiv preprint arXiv:1809.06683*.

Grainger, Jonathan & Dijkstra, Ton (1992). “On the representation and use of language information in bilinguals”. *Advances in psychology*. Vol. 83. Elsevier, pp. 207–220.

Gravanis, Georgios, Vakali, Athena, Diamantaras, Konstantinos, & Karadais, Panagiotis (2019). “Behind the cues: A benchmarking study for fake news detection”. *Expert Systems with Applications* 128, pp. 201–213.

Greff, Klaus, Srivastava, Rupesh K, Koutník, Jan, Steunebrink, Bas R, & Schmidhuber, Jürgen (2016). “LSTM: A search space odyssey”. *IEEE transactions on neural networks and learning systems* 28.10, pp. 2222–2232.

Guo, Bao, Zhang, Chunxia, Liu, Junmin, & Ma, Xiaoyi (2019). “Improving text classification with weighted word embeddings via a multi-channel TextCNN model”. *Neurocomputing* 363, pp. 366–374.

Gupta, Manish, Zhao, Peixiang, & Han, Jiawei (2012). “Evaluating event credibility on twitter”. *Proceedings of the 2012 SIAM International Conference on Data Mining*. SIAM, pp. 153–164.

Gupta, Shashank, Thirukovalluru, Raghuveer, Sinha, Manjira, & Mannarswamy, Sandya (2018). “CIMTDetect: A Community Infused Matrix-Tensor Coupled Factorization Based Method for Fake News Detection”. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, pp. 278–281.

H. De Vreese Jochen Peter, Holli A. Semetko Claes (2001). “Framing politics at the launch of the Euro: A cross-national comparative study of frames in the news”. *Political communication* 18.2, pp. 107–122.

Hailong, Zhang, Wenyan, Gan, & Bo, Jiang (2014). “Machine learning and lexicon based methods for sentiment classification: A survey”. *2014 11th Web Information System and Application Conference*. IEEE, pp. 262–265.

Han, Jingyu, Chen, Kejia, & Jiang, Dawei (2012). “Probabilistically ranking web article quality based on evolution patterns”. *Transactions on Large-Scale Data-and Knowledge-Centered Systems VI*. Springer, pp. 229–258.

Harshman, Richard A et al. (1970). "Foundations of the PARAFAC procedure: Models and conditions for an" explanatory" multimodal factor analysis".

He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing, & Sun, Jian (2015). "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification". *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034.

Hochreiter, Sepp & Schmidhuber, Jürgen (1997). "Long short-term memory". *Neural computation* 9.8, pp. 1735–1780.

Hosmer Jr, David W, Lemeshow, Stanley, & Sturdivant, Rodney X (2013). *Applied logistic regression*. Vol. 398. John Wiley & Sons.

Hosseinimotlagh, Seyedmehdi & Papalexakis, Evangelos E (2018). "Unsupervised content-based identification of fake news articles with tensor decomposition ensembles". *MIS2, Marina Del Rey, CA, USA*.

Huang, Zhiheng, Xu, Wei, & Yu, Kai (2015). "Bidirectional LSTM-CRF models for sequence tagging". *arXiv preprint arXiv:1508.01991*.

Ishida, Yoshiteru & Kuraya, Sanae (2018). "Fake News and its Credibility Evaluation by Dynamic Relational Networks: A Bottom up Approach". *Procedia computer science* 126, pp. 2228–2237.

Iyyer, Mohit, Manjunatha, Varun, Boyd-Graber, Jordan, & Daumé III, Hal (2015). "Deep unordered composition rivals syntactic methods for text classification". *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Vol. 1, pp. 1681–1691.

Jain, Anil K, Mao, Jianchang, & Mohiuddin, K Moidin (1996). "Artificial neural networks: A tutorial". *Computer* 29.3, pp. 31–44.

Joachims, Thorsten (1998). "Text categorization with support vector machines: Learning with many relevant features". *European conference on machine learning*. Springer, pp. 137–142.

Johnson, David Edward (1998). *Multimodal natural language interface for cross-application tasks*. US Patent 5,748,974.

Jwa, Heejung, Oh, Dongsuk, Park, Kinam, Kang, Jang Mook, & Lim, Heuiseok (2019). "exBAKE: automatic fake news detection model based on bidirectional encoder representations from transformers (bert)". *Applied Sciences* 9.19, p. 4062.

Kalchbrenner, Nal, Grefenstette, Edward, & Blunsom, Phil (2014). "A Convolutional Neural Network for Modelling Sentences". *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 655–665.

Kaliyar, Rohit Kumar, Goswami, Anurag, & Narang, Pratik (2020a). "DeepFakE: improving fake news detection using tensor decomposition-based deep neural network". *The Journal of Supercomputing*, pp. 1–23.

Kaliyar, Rohit Kumar, Goswami, Anurag, & Narang, Pratik (2021a). "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach". *Multimedia Tools and Applications*, pp. 1–24.

Kaliyar, Rohit Kumar, Goswami, Anurag, & Narang, Pratik (2021b). "MCNNet: Generalizing Fake News Detection with a Multichannel Convolutional Neural Network using a Novel COVID-19 Dataset". *8th ACM IKDD CODS and 26th COMAD*, pp. 437–437.

Kaliyar, Rohit Kumar, Goswami, Anurag, Narang, Pratik, & Sinha, Soumendu (2020b). "FNDNet–a deep convolutional neural network for fake news detection". *Cognitive Systems Research* 61, pp. 32–44.

Kamkarhaghighi, Mehran & Makrehchi, Masoud (2017). "Content tree word embedding for document representation". *Expert Systems with Applications* 90, pp. 241–249.

Karimi, Hamid, Roy, Proteek, Saba-Sadiya, Sari, & Tang, Jiliang (2018). "Multi-source multi-class fake news detection". *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 1546–1557.

Khatri, CG & Rao, C Radhakrishna (1968). “Solutions to some functional equations and their applications to characterization of probability distributions”. *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 167–180.

Kim, Jooyeon, Tabibian, Behzad, Oh, Alice, Schölkopf, Bernhard, & Gomez-Rodriguez, Manuel (2018). “Leveraging the crowd to detect and reduce the spread of fake news and misinformation”. *Proceedings of the eleventh ACM international conference on web search and data mining*, pp. 324–332.

Kim, Yoon (2014). “Convolutional neural networks for sentence classification”. *arXiv preprint arXiv:1408.5882*.

Kim, Yoon, Jernite, Yacine, Sontag, David, & Rush, Alexander M (2016). “Character-aware neural language models”. *Thirtieth AAAI Conference on Artificial Intelligence*.

Kingma, Diederik P & Ba, Jimmy (2014). “Adam: A method for stochastic optimization”. *arXiv preprint arXiv:1412.6980*.

Kumar, KP Krishna & Geethakumari, G (2014). “Detecting misinformation in online social networks using cognitive psychology”. *Human-centric Computing and Information Sciences* 4.1, p. 14.

Kumar, Sachin, Asthana, Rohan, Upadhyay, Shashwat, Upreti, Nidhi, & Akbar, Mohammad (2020). “Fake news detection using deep learning models: A novel approach”. *Transactions on Emerging Telecommunications Technologies* 31.2, e3767.

Kumar, Srijan & Shah, Neil (2018). “False information on web and social media: A survey”. *arXiv preprint arXiv:1804.08559*.

Kusner, Matt, Sun, Yu, Kolkin, Nicholas, & Weinberger, Kilian (2015). “From word embeddings to document distances”. *International conference on machine learning*, pp. 957–966.

Lebret, Rémi & Collobert, Ronan (2015). “Rehabilitation of count-based models for word vector representations”. *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer, pp. 417–429.

LeCun, Yann, Bengio, Yoshua, & Hinton, Geoffrey (2015). "Deep learning". *nature* 521.7553, p. 436.

Lee, Daniel D & Seung, H Sebastian (2001). "Algorithms for non-negative matrix factorization". *Advances in neural information processing systems*, pp. 556–562.

Lee, Jinhyuk, Yoon, Wonjin, Kim, Sungdong, Kim, Donghyeon, Kim, Sunkyu, So, Chan Ho, & Kang, Jaewoo (2019). "Biobert: pre-trained biomedical language representation model for biomedical text mining". *arXiv preprint arXiv:1901.08746*.

Li, Yuanzhi & Yuan, Yang (2017). "Convergence analysis of two-layer neural networks with relu activation". *Advances in Neural Information Processing Systems*, pp. 597–607.

Liang, Gang, He, Wenbo, Xu, Chun, Chen, Liangyin, & Zeng, Jinquan (2015). "Rumor identification in microblogging systems based on users' behavior". *IEEE Transactions on Computational Social Systems* 2.3, pp. 99–108.

Liao, Yihua & Vemuri, V Rao (2002). "Use of k-nearest neighbor classifier for intrusion detection". *Computers & security* 21.5, pp. 439–448.

Liaw, Andy, Wiener, Matthew, et al. (2002). "Classification and regression by randomForest". *R news* 2.3, pp. 18–22.

Liu, Yang & Wu, Yi-Fang Brook (2018). "Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks". *Thirty-Second AAAI Conference on Artificial Intelligence*.

Long, Yunfei, Lu, Qin, Xiang, Rong, Li, Minglei, & Huang, Chu-Ren (2017). "Fake news detection through multi-perspective speaker profiles". *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pp. 252–256.

Ma, Jing, Gao, Wei, Mitra, Prasenjit, Kwon, Sejeong, Jansen, Bernard J, Wong, Kam-Fai, & Cha, Meeyoung (2016). "Detecting rumors from microblogs with recurrent neural networks." *Ijcai*, pp. 3818–3824.

Ma, Jing, Gao, Wei, & Wong, Kam-Fai (2018). "Rumor Detection on Twitter with Tree-structured Recursive Neural Networks". *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1980–1989.

MacKay, David JC (1999). "Comparison of approximate methods for handling hyperparameters". *Neural computation* 11.5, pp. 1035–1068.

Malik, Sharad, Sentovich, Ellen M, Brayton, Robert K, & Sangiovanni-Vincentelli, Alberto (1991). "Retiming and resynthesis: Optimizing sequential networks with combinational techniques". *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 10.1, pp. 74–84.

Manning, Christopher D (2015). "Computational linguistics and deep learning". *Computational Linguistics* 41.4, pp. 701–707.

McBeth, Mark K & Clemons, Randy S (2011). "Is fake news the real news". *The Stewart/Colbert effect: Essays on the real impacts of fake news*, pp. 79–98.

Medhat, Walaa, Hassan, Ahmed, & Korashy, Hoda (2014). "Sentiment analysis algorithms and applications: A survey". *Ain Shams engineering journal* 5.4, pp. 1093–1113.

Mikolov, Tomas, Chen, Kai, Corrado, Greg, & Dean, Jeffrey (2013a). "Efficient estimation of word representations in vector space". *arXiv preprint arXiv:1301.3781*.

Mikolov, Tomas, Grave, Edouard, Bojanowski, Piotr, Puhrsch, Christian, & Joulin, Armand (2017). "Advances in pre-training distributed word representations". *arXiv preprint arXiv:1712.09405*.

Mikolov, Tomas, Sutskever, Ilya, Chen, Kai, Corrado, Greg S, & Dean, Jeff (2013b). "Distributed representations of words and phrases and their compositionality". *Advances in neural information processing systems*, pp. 3111–3119.

Miller, Thomas P & Leon, Adele (2017). "Introduction to special issue on literacy, democracy, and fake news: Making it right in the era of fast and slow literacies". *Literacy in Composition Studies* 5.2, pp. 10–23.

Mitra, Tanushree & Gilbert, Eric (2015). “Credbank: A large-scale social media corpus with associated credibility annotations”. *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 9. 1.

Monteiro, Rafael A, Santos, Roney LS, Pardo, Thiago AS, Almeida, Tiago A de, Ruiz, Evandro ES, & Vale, Oto A (2018). “Contributions to the Study of Fake News in Portuguese: New Corpus and Automatic Detection Results”. *International Conference on Computational Processing of the Portuguese Language*. Springer, pp. 324–334.

Moreno, Pedro J, Logan, Beth, & Raj, Bhiksha (2001). “A boosting approach for confidence scoring”. *Seventh European Conference on Speech Communication and Technology*.

Mou, Lili, Peng, Hao, Li, Ge, Xu, Yan, Zhang, Lu, & Jin, Zhi (2015). “Discriminative neural sentence modeling by tree-based convolution”. *arXiv preprint arXiv:1504.01106*.

Munandar, Devi, Arisal, Andria, Riswantini, Dianadewi, & Rozie, Andri Fachrur (2018). “Text classification for sentiment prediction of social media dataset using multichannel convolution neural network”. *2018 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*. IEEE, pp. 104–109.

Nagi, Jawad, Ducatelle, Frederick, Di Caro, Gianni A, Cireşan, Dan, Meier, Ueli, Giusti, Alessandro, Nagi, Farrukh, Schmidhuber, Jürgen, & Gambardella, Luca Maria (2011). “Max-pooling convolutional neural networks for vision-based hand gesture recognition”. *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. IEEE, pp. 342–347.

Natekin, Alexey & Knoll, Alois (2013). “Gradient boosting machines, a tutorial”. *Frontiers in neurorobotics* 7, p. 21.

O’Brien, Nicole, Latessa, Sophia, Evangelopoulos, Georgios, & Boix, Xavier (2018). “The language of fake news: Opening the black-box of deep learning based detectors”.

Oliveira, Nicollas R de, Pisa, Pedro S, Lopez, Martin Andreoni, Medeiros, Dianne Scherly V de, & Mattos, Diogo MF (2021). “Identifying Fake News on Social Networks Based on Natural Language Processing: Trends and Challenges”. *Information* 12.1, p. 38.

Orso, Daniele, Federici, Nicola, Copetti, Roberto, Vetrugno, Luigi, & Bove, Tiziana (2020). "Infodemic and the spread of fake news in the COVID-19-era". *European Journal of Emergency Medicine*.

Oshikawa, Ray, Qian, Jing, & Wang, William Yang (2020). "A Survey on Natural Language Processing for Fake News Detection". *Proceedings of The 12th Language Resources and Evaluation Conference*, pp. 6086–6093.

Ott, Myle, Choi, Yejin, Cardie, Claire, & Hancock, Jeffrey T (2011). "Finding deceptive opinion spam by any stretch of the imagination". *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1*. Association for Computational Linguistics, pp. 309–319.

Ozbay, Feyza Altunbey & Alatas, Bilal (2020). "Fake news detection within online social media using supervised artificial intelligence algorithms". *Physica A: Statistical Mechanics and its Applications* 540, p. 123174.

Paisitkriangkrai, Sakrapee, Sherrah, Jamie, Janney, Pranam, Hengel, Van-Den, et al. (2015). "Effective semantic pixel labelling with convolutional networks and conditional random fields". *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 36–43.

Pal, Mahesh (2005). "Random forest classifier for remote sensing classification". *International Journal of Remote Sensing* 26.1, pp. 217–222.

Pan, Jeff Z, Pavlova, Siyana, Li, Chenxi, Li, Ningxi, Li, Yangmei, & Liu, Jinshuo (2018). "Content based fake news detection using knowledge graphs". *International Semantic Web Conference*. Springer, pp. 669–683.

Papanastasiou, Frosso, Katsimpras, Georgios, & Paliouras, Georgios (2019). "Tensor Factorization with Label Information for Fake News Detection". *arXiv preprint arXiv:1908.03957*.

Patidar, Raghavendra, Sharma, Lokesh, et al. (2011). "Credit card fraud detection using neural network". *International Journal of Soft Computing and Engineering (IJSCE)* 1.32-38.

Pennington, Jeffrey, Socher, Richard, & Manning, Christopher (2014). "Glove: Global vectors for word representation". *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543.

Pérez-Rosas, Verónica, Kleinberg, Bennett, Lefevre, Alexandra, & Mihalcea, Rada (2017). "Automatic detection of fake news". *arXiv preprint arXiv:1708.07104*.

Persily, Nathaniel (2017). "The 2016 US Election: Can democracy survive the internet?" *Journal of democracy* 28.2, pp. 63–76.

Peters, Matthew E, Neumann, Mark, Iyyer, Mohit, Gardner, Matt, Clark, Christopher, Lee, Kenton, & Zettlemoyer, Luke (2018). "Deep contextualized word representations". *arXiv preprint arXiv:1802.05365*.

Pierri, Francesco & Ceri, Stefano (2019). "False news on social media: a data-driven survey". *ACM Sigmod Record* 48.2, pp. 18–27.

Popat, Kashyap, Mukherjee, Subhabrata, Yates, Andrew, & Weikum, Gerhard (2018). "DeClarE: Debunking fake news and false claims using evidence-aware deep learning". *arXiv preprint arXiv:1809.06416*.

Potthast, Martin, Kiesel, Johannes, Reinartz, Kevin, Bevendorff, Janek, & Stein, Benno (2017). "A stylometric inquiry into hyperpartisan and fake news". *arXiv preprint arXiv:1702.05638*.

Qassim, Hussam, Verma, Abhishek, & Feinzimer, David (2018). "Compressed residual-VGG16 CNN model for big data places image recognition". *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, pp. 169–175.

Qi, Ye, Sachan, Devendra Singh, Felix, Matthieu, Padmanabhan, Sarguna Janani, & Neubig, Graham (2018). "When and why are pre-trained word embeddings useful for neural machine translation?" *arXiv preprint arXiv:1804.06323*.

Rabanser, Stephan, Shchur, Oleksandr, & Günnemann, Stephan (2017). "Introduction to tensor decompositions and their applications in machine learning". *arXiv preprint arXiv:1711.10781*.

Rashkin, Hannah, Choi, Eunsol, Jang, Jin Yea, Volkova, Svitlana, & Choi, Yejin (2017). "Truth of varying shades: Analyzing language in fake news and political fact-checking". *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 2931–2937.

Ren, Yafeng, Wang, Ruimin, & Ji, Donghong (2016). "A topic-enhanced word embedding for Twitter sentiment classification". *Information Sciences* 369, pp. 188–198.

Rish, Irina et al. (2001). "An empirical study of the naive Bayes classifier". *IJCAI 2001 workshop on empirical methods in artificial intelligence*. Vol. 3. 22, pp. 41–46.

ROSSI, GUSTAF (n.d.). "Real news or Fake News" ().

Roy, Arjun, Basak, Kingshuk, Ekbal, Asif, & Bhattacharyya, Pushpak (2018). "A Deep Ensemble Framework for Fake News Detection and Classification". *arXiv preprint arXiv:1811.04670*.

Rubin, Victoria L, Conroy, Niall, Chen, Yimin, & Cornwell, Sarah (2016). "Fake news or truth? using satirical cues to detect potentially misleading news". *Proceedings of the second workshop on computational approaches to deception detection*, pp. 7–17.

Ruchansky, Natali, Seo, Sungyong, & Liu, Yan (2017). "Csi: A hybrid deep model for fake news detection". *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, pp. 797–806.

Safavian, S Rasoul & Landgrebe, David (1991). "A survey of decision tree classifier methodology". *IEEE transactions on systems, man, and cybernetics* 21.3, pp. 660–674.

Sainath, Tara N, Vinyals, Oriol, Senior, Andrew, & Sak, Haşim (2015). "Convolutional, long short-term memory, fully connected deep neural networks". *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, pp. 4580–4584.

Salem, Fatima K Abu, Al Feel, Roaa, Elbassuoni, Shady, Jaber, Mohamad, & Farah, May (2019). "Fa-kes: A fake news dataset around the syrian war". *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 13, pp. 573–582.

Santia, Giovanni & Williams, Jake (2018). “Buzzface: A news veracity dataset with facebook user commentary and egos”. *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 12. 1.

Schmidhuber, Jürgen (2015). “Deep learning in neural networks: An overview”. *Neural networks* 61, pp. 85–117.

Sebastiani, Fabrizio (2002). “Machine learning in automated text categorization”. *ACM computing surveys (CSUR)* 34.1, pp. 1–47.

Seide, Frank, Li, Gang, Chen, Xie, & Yu, Dong (2011). “Feature engineering in context-dependent deep neural networks for conversational speech transcription”. *2011 IEEE Workshop on Automatic Speech Recognition & Understanding*. IEEE, pp. 24–29.

Shao, Chengcheng, Ciampaglia, Giovanni Luca, Varol, Onur, Flammini, Alessandro, & Menczer, Filippo (2017). “The spread of fake news by social bots”. *arXiv preprint arXiv:1707.07592* 96, p. 104.

Sharma, Karishma, Qian, Feng, Jiang, He, Ruchansky, Natali, Zhang, Ming, & Liu, Yan (2019). “Combating fake news: A survey on identification and mitigation techniques”. *ACM Transactions on Intelligent Systems and Technology (TIST)* 10.3, pp. 1–42.

Shin, Hoo-Chang, Roth, Holger R, Gao, Mingchen, Lu, Le, Xu, Ziyue, Nogues, Isabella, Yao, Jianhua, Mollura, Daniel, & Summers, Ronald M (2016). “Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning”. *IEEE transactions on medical imaging* 35.5, pp. 1285–1298.

Shin, Jieun, Jian, Lian, Driscoll, Kevin, & Bar, François (2018). “The diffusion of misinformation on social media: Temporal pattern, message, and source”. *Computers in Human Behavior* 83, pp. 278–287.

Shu, Kai, Bernard, H Russell, & Liu, Huan (2019a). “Studying fake news via network analysis: detection and mitigation”. *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining*. Springer, pp. 43–65.

Shu, Kai, Cui, Limeng, Wang, Suhang, Lee, Dongwon, & Liu, Huan (2019b). "defend: Explainable fake news detection". *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 395–405.

Shu, Kai, Mahudeswaran, Deepak, & Liu, Huan (2019c). "FakeNewsTracker: a tool for fake news collection, detection, and visualization". *Computational and Mathematical Organization Theory* 25.1, pp. 60–71.

Shu, Kai, Mahudeswaran, Deepak, Wang, Suhang, Lee, Dongwon, & Liu, Huan (2018). "Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media". *arXiv preprint arXiv:1809.01286*.

Shu, Kai, Sliva, Amy, Wang, Suhang, Tang, Jiliang, & Liu, Huan (2017). "Fake news detection on social media: A data mining perspective". *ACM SIGKDD Explorations Newsletter* 19.1, pp. 22–36.

Shu, Kai, Wang, Suhang, & Liu, Huan (2019d). "Beyond news contents: The role of social context for fake news detection". *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. ACM, pp. 312–320.

Sibi, P, Jones, S Allwyn, & Siddarth, P (2013). "Analysis of different activation functions using back propagation neural networks". *Journal of Theoretical and Applied Information Technology* 47.3, pp. 1264–1268.

Sicilia, Rosa, Giudice, Stella Lo, Pei, Yulong, Pechenizkiy, Mykola, & Soda, Paolo (2018). "Twitter rumour detection in the health domain". *Expert Systems with Applications* 110, pp. 33–40.

Singh, DSKR Vivek, Dasgupta, Rupanjal, & Ghosh, I (2017). "Automated fake news detection using linguistic analysis and machine learning". *International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation (SBP-BRiMS)*, pp. 1–3.

Singhania, Sneha, Fernandez, Nigel, & Rao, Shrisha (2017). "3han: A deep neural network for fake news detection". *International Conference on Neural Information Processing*. Springer, pp. 572–581.

Srivastava, Nitish, Hinton, Geoffrey, Krizhevsky, Alex, Sutskever, Ilya, & Salakhutdinov, Ruslan (2014). “Dropout: a simple way to prevent neural networks from overfitting”. *The journal of machine learning research* 15.1, pp. 1929–1958.

Stein, Roger Alan, Jaques, Patricia A, & Valiati, João Francisco (2019). “An analysis of hierarchical text classification using word embeddings”. *Information Sciences* 471, pp. 216–232.

Stringhini, Gianluca, Kruegel, Christopher, & Vigna, Giovanni (2010). “Detecting spammers on social networks”. *Proceedings of the 26th annual computer security applications conference*, pp. 1–9.

Su, Qi, Wan, Mingyu, Liu, Xiaoqian, & Huang, Chu-Ren (2020). “Motivations, Methods and Metrics of Misinformation Detection: An NLP Perspective”. *Natural Language Processing Research* 1, pp. 1–13.

Tacchini, Eugenio, Ballarin, Gabriele, Della Vedova, Marco L, Moret, Stefano, & Alfaro, Luca de (2017). “Some like it hoax: Automated fake news detection in social networks”. *arXiv preprint arXiv:1704.07506*.

Tai, Kai Sheng, Socher, Richard, & Manning, Christopher D (2015). “Improved semantic representations from tree-structured long short-term memory networks”. *arXiv preprint arXiv:1503.00075*.

Tang, Duyu, Wei, Furu, Yang, Nan, Zhou, Ming, Liu, Ting, & Qin, Bing (2014). “Learning sentiment-specific word embedding for twitter sentiment classification”. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1555–1565.

Tenney, Ian, Das, Dipanjan, & Pavlick, Ellie (2019). “Bert rediscovers the classical nlp pipeline”. *arXiv preprint arXiv:1905.05950*.

Thota, Aswini, Tilak, Priyanka, Ahluwalia, Simrat, & Lohia, Nibrat (2018). “Fake News Detection: A Deep Learning Approach”. *SMU Data Science Review* 1.3, p. 10.

Torlay, L, Perrone-Bertolotti, Marcela, Thomas, Elizabeth, & Baciu, Monica (2017). “Machine learning–XGBoost analysis of language networks to classify patients with epilepsy”. *Brain informatics* 4.3, p. 159.

Vasudevan, Vijay, Zoph, Barret, Shlens, Jonathon, & Le, Quoc V (2019). *Neural architecture search for convolutional neural networks*. US Patent App. 16/040,067.

Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, Uszkoreit, Jakob, Jones, Llion, Gomez, Aidan N, Kaiser, Łukasz, & Polosukhin, Illia (2017). "Attention is all you need". *Advances in neural information processing systems*, pp. 5998–6008.

Vicario, Michela Del, Quattrociocchi, Walter, Scala, Antonio, & Zollo, Fabiana (2019). "Polarization and fake news: Early warning of potential misinformation targets". *ACM Transactions on the Web (TWEB)* 13.2, pp. 1–22.

Vishwakarma, Dinesh Kumar, Varshney, Deepika, & Yadav, Ashima (2019). "Detection and veracity analysis of fake news via scrapping and authenticating the web search". *Cognitive Systems Research* 58, pp. 217–229.

Vosoughi, Soroush, Mohsenvand, Mostafa'Neo', & Roy, Deb (2017). "Rumor gauge: predicting the veracity of rumors on twitter". *ACM Transactions on Knowledge Discovery from Data (TKDD)* 11.4, p. 50.

Wager, Stefan, Wang, Sida, & Liang, Percy (2013). "Dropout training as adaptive regularization". *arXiv preprint arXiv:1307.1493*.

Wakita, Ken & Tsurumi, Toshiyuki (2007). "Finding community structure in mega-scale social networks". *Proceedings of the 16th international conference on World Wide Web*, pp. 1275–1276.

Wang, William Yang (2017). ""liar, liar pants on fire": A new benchmark dataset for fake news detection". *In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* 2.1, pp. 422–426.

Wang, Yaqing, Ma, Fenglong, Jin, Zhiwei, Yuan, Ye, Xun, Guangxu, Jha, Kishlay, Su, Lu, & Gao, Jing (2018). "Eann: Event adversarial neural networks for multi-modal fake news detection". *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*. ACM, pp. 849–857.

Wang, Yequan, Huang, Minlie, Zhao, Li, et al. (2016). "Attention-based LSTM for aspect-level sentiment classification". *Proceedings of the 2016 conference on empirical methods in natural language processing*, pp. 606–615.

Wang, Yuhang, Wang, Li, Yang, Yanjie, & Lian, Tao (2021). "SemSeq4FD: Integrating global semantic relationship and local sequential order to enhance text representation for fake news detection". *Expert Systems with Applications* 166, p. 114090.

Wason, Ritika (2018). "Deep learning: Evolution and expansion". *Cognitive Systems Research* 52, pp. 701–708.

Wei, Wei, Zhou, Bin, Maskeliūnas, Rytis, Damaševičius, Robertas, Połap, Dawid, & Woźniak, Marcin (2019). "Iterative Design and Implementation of Rapid Gradient Descent Method". *International Conference on Artificial Intelligence and Soft Computing*. Springer, pp. 530–539.

Weir, William (2009). *History's Greatest Lies: The Startling Truths Behind World Events Our History Books Got Wrong*. Fair Winds Press.

Weiss, Andrew P, Alwan, Ahmed, Garcia, Eric P, & Garcia, Julieta (2020). "Surveying fake news: Assessing university faculty's fragmented definition of fake news and its impact on teaching critical thinking". *International Journal for Educational Integrity* 16.1, pp. 1–30.

Wu, Haibing & Gu, Xiaodong (2015). "Max-pooling dropout for regularization of convolutional neural networks". *International Conference on Neural Information Processing*. Springer, pp. 46–54.

Wu, Liang, Li, Jundong, Hu, Xia, & Liu, Huan (2017). "Gleaning wisdom from the past: Early detection of emerging rumors in social media". *Proceedings of the 2017 SIAM international conference on data mining*. SIAM, pp. 99–107.

Xu, Kelvin, Ba, Jimmy, Kiros, Ryan, Cho, Kyunghyun, Courville, Aaron, Salakhudinov, Ruslan, Zemel, Rich, & Bengio, Yoshua (2015). "Show, attend and tell: Neural image caption generation with visual attention". *International conference on machine learning*, pp. 2048–2057.

Yang, Fan, Liu, Yang, Yu, Xiaohui, & Yang, Min (2012). “Automatic detection of rumor on Sina Weibo”. *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics*. ACM, p. 13.

Yang, Shuo, Shu, Kai, Wang, Suhang, Gu, Renjie, Wu, Fan, & Liu, Huan (2019). “Unsupervised fake news detection on social media: A generative approach”. *Proceedings of 33rd AAAI Conference on Artificial Intelligence*.

Yang, Yang, Zheng, Lei, Zhang, Jiawei, Cui, Qingcai, Li, Zhoujun, & Yu, Philip S (2018). “TI-CNN: Convolutional neural networks for fake news detection”. *arXiv preprint arXiv:1806.00749*.

Young, Tom, Hazarika, Devamanyu, Poria, Soujanya, & Cambria, Erik (2018). “Recent trends in deep learning based natural language processing”. *ieee Computational intelligenCe magazine* 13.3, pp. 55–75.

Yu, Feng, Liu, Qiang, Wu, Shu, Wang, Liang, & Tan, Tieniu (2017). “A convolutional approach for misinformation identification”. *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pp. 3901–3907.

Zagoruyko, Sergey & Komodakis, Nikos (2015). “Learning to compare image patches via convolutional neural networks”. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4353–4361.

Zhang, Jiawei, Cui, Limeng, Fu, Yanjie, & Gouza, Fisher B (2018). “Fake news detection with deep diffusive network model”. *arXiv preprint arXiv:1805.08751*.

Zhang, Jiawei, Dong, Bowen, & Philip, S Yu (2020). “Fakedetector: Effective fake news detection with deep diffusive neural network”. *2020 IEEE 36th International Conference on Data Engineering (ICDE)*. IEEE, pp. 1826–1829.

Zhang, Qishan, Qiu, Qirong, Guo, Wenzhong, Guo, Kun, & Xiong, Naixue (2016). “A social community detection algorithm based on parallel grey label propagation”. *Computer Networks* 107, pp. 133–143.

Zhang, Xiang, Zhao, Junbo, & LeCun, Yann (2015). “Character-level convolutional networks for text classification”. *Advances in neural information processing systems*, pp. 649–657.

Zhang, Xichen & Ghorbani, Ali A (2020). "An overview of online fake news: Characterization, detection, and discussion". *Information Processing & Management* 57.2, p. 102025.

Zhang, Y & Wallace, BC (2016). "A sensitivity analysis of convolutional neural networks for sentence classification". *arXiv preprint arXiv:1510.03820*.

Zhang, Ye & Wallace, Byron (2015). "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification". *arXiv preprint arXiv:1510.03820*.

Zhao, Zhe, Resnick, Paul, & Mei, Qiaozhu (2015). "Enquiring minds: Early detection of rumors in social media from enquiry posts". *Proceedings of the 24th international conference on world wide web*, pp. 1395–1405.

Zheng, Xianghan, Zeng, Zhipeng, Chen, Zheyi, Yu, Yuanlong, & Rong, Chunming (2015). "Detecting spammers on social networks". *Neurocomputing* 159, pp. 27–34.

Zhong, Botao, Xing, Xuejiao, Love, Peter, Wang, Xu, & Luo, Hanbin (2019). "Convolutional neural network: Deep learning-based classification of building quality problems". *Advanced Engineering Informatics* 40, pp. 46–57.

Zhong, Shangping, Chen, Tianshun, He, Fengying, & Niu, Yuzhen (2014). "Fast Gaussian kernel learning for classification tasks based on specially structured global optimization". *Neural networks* 57, pp. 51–62.

Zhou, Chunting, Sun, Chonglin, Liu, Zhiyuan, & Lau, Francis (2015). "A C-LSTM neural network for text classification". *arXiv preprint arXiv:1511.08630*.

Zhou, Xinyi, Wu, Jindi, & Zafarani, Reza (2020). "Similarity Aware Multi-modal Fake News Detection". *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, pp. 354–367.

Zhou, Xinyi & Zafarani, Reza (2018). "Fake news: A survey of research, detection methods, and opportunities". *arXiv preprint arXiv:1812.00315*.

Zhou, Xinyi & Zafarani, Reza (2019). "Network-based fake news detection: A pattern-driven approach". *ACM SIGKDD Explorations Newsletter* 21.2, pp. 48–60.

Zubiaga, Arkaitz, Aker, Ahmet, Bontcheva, Kalina, Liakata, Maria, & Procter, Rob (2018a). "Detection and resolution of rumours in social media: A survey". *ACM Computing Surveys (CSUR)* 51.2, p. 32.

Zubiaga, Arkaitz, Kochkina, Elena, Liakata, Maria, Procter, Rob, Lukasik, Michal, Bontcheva, Kalina, Cohn, Trevor, & Augenstein, Isabelle (2018b). "Discourse-aware rumour stance classification in social media using sequential classifiers". *Information Processing & Management* 54.2, pp. 273–290.

Zubiaga, Arkaitz, Liakata, Maria, & Procter, Rob (2016). "Learning Reporting Dynamics during Breaking News for Rumour Detection in Social Media". *arXiv*, arXiv–1610.

Zurada, Jacek M (1992). *Introduction to artificial neural systems*. Vol. 8. West publishing company St. Paul.

Publications

SCI Journal Papers

Published

- **Kaliyar, R. K.**, Goswami, A., Narang, P., & Sinha, S. (2020). FNDNet–a deep convolutional neural network for fake news detection. Cognitive Systems Research, 61, 32-44. https://doi.org/10.1016/j.cogsys.2019.12.005. **(Elsevier, SCI Impact Factor: 1.902).**

- **Kaliyar, R. K.**, Goswami, A., & Narang, P. (2020). DeepFakE: improving fake news detection using tensor decomposition-based deep neural network. The Journal of Supercomputing, 1-23. https://doi.org/10.1007/s11227-020-03294-y. **(Springer Nature, SCI Impact Factor: 2.469).**

- **Kaliyar, R. K.**, Goswami, A., & Narang, P. (2021) EchoFakeD: improving fake news detection in social media with an efficient deep neural network. Neural Computing and Applications, 1-17. https://doi.org/10.1007/s00521-020-05611-1. **(Springer Nature, SCI Impact Factor: 4.774).**

- **Kaliyar, R. K.**, Goswami, A., & Narang, P. (2021) FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. Multimedia Tools and Applications, 1-24. https://doi.org/10.1007/s11042-020-10183-2. **(Springer Nature, SCI Impact Factor: 2.313).**

In Press / Under Review

- Jain, V., **Kaliyar, R. K.**, Goswami, A., Narang, P. AENeT: An Attention Enabled Neural Architecture for Fake News Detection using Contextual Features. Neural Computing and Applications, **(Elsevier, SCI Impact Factor: 4.774).**

- **Kaliyar, R. K.**, Goswami, A., Narang, P. Can Deep Neural Networks Learn to Generalize Fake News Detection?. IEEE-IT Professional,
(IEEE, SCI Impact Factor: 3.700).

Conference Papers

- **R. K. Kaliyar**, A. Goswami and P. Narang, "Multiclass Fake News Detection using Ensemble Machine Learning," 2019 IEEE 9th International Conference on Advanced Computing (IACC), Tiruchirappalli, India, 2019, pp. 103-107,
doi: 10.1109/IACC48062.2019.8971579.

- **Kaliyar, R. K.**, Goswami, A., & Narang, P. (2021). MCNNet: Generalizing Fake News Detection with a Multichannel Convolutional Neural Network using a Novel COVID-19 Dataset. In 8th ACM IKDD CODS and 26th COMAD (pp. 437-437).

- **Kaliyar, R.K.;** Goswami, A. and Narang, P. (2021). A Hybrid Model for Effective Fake News Detection with a Novel COVID-19 Dataset.In Proceedings of the 13th International Conference on Agents and Artificial Intelligence - Volume 2: ICAART, ISBN 978-989-758-484-8, pages 1066-1072. DOI: 10.5220/0010316010661072