

Human Reliability Analysis using AI



By

Muhammad Abdullah

(Registration No: 00000327296)

Department of Robotics and Artificial Intelligence

School of Mechanical and Manufacturing Engineering

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(2024)

Human Reliability Analysis using AI



By

Muhammad Abdullah

(Registration No: 00000327296)

A thesis submitted to the National University of Sciences and Technology, Islamabad,

in partial fulfillment of the requirements for the degree of

Master of Science in
Robotics and Intelligent Machine Engineering

Supervisor: Dr. Yasar Ayaz

School of Mechanical and Manufacturing Engineering

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(2024)

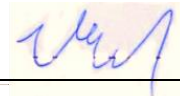
THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS thesis written by **Mr. Muhammad Abdullah**, (Registration No. **00000327296**), of **SCHOOL OF MECHANICAL & MANUFACTURING ENGINEERING (SMME)** has been vetted by undersigned, found complete in all respects as per NUST Statutes/ Regulations/ Masters Policy, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of Masters degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis titled **“Human Reliability Analysis using AI”**.


Signature:  _____

Name of Supervisor: Dr. Yasar Ayaz

Date: 30-09-2024

Signature (HOD):  _____

Date: 30-09-2024

Signature (Dean):  _____


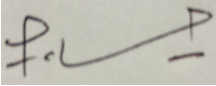
Date: 30-09-2024




National University of Sciences & Technology (NUST)
MASTER'S THESIS WORK

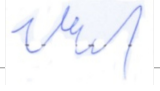
We hereby recommend that the dissertation prepared under our supervision by: Muhammad Abdullah (00000327296)
Titled: Human Reliability Analysis Using AI be accepted in partial fulfillment of the requirements for the award of MS in Robotics & Intelligent Machine Engineering degree.

Examination Committee Members

- | | | |
|----|---------------------------|--|
| 1. | Name: Umer Asgher | Signature:  |
| 2. | Name: Khawaja Fahad Iqbal | Signature:  |

Supervisor: Yasar Ayaz

Signature: 
Date: 30 - Sep - 2024



Head of Department

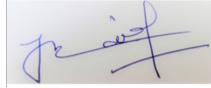
30 - Sep - 2024

Date

COUNTERSIGNED

30 - Sep - 2024

Date



Dean/Principal

CERTIFICATE OF APPROVAL

This is to certify that the research work presented in this thesis, entitled “**Human Reliability Analysis using AI**” was conducted by **Mr. Muhammad Abdullah** under the supervision of **Dr. Yasar Ayaz**.

No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to the **Department of Robotics & Artificial Intelligence** in partial fulfillment of the requirements for the degree of Master of Science in the Field of **Robotics and Intelligent Machine Engineering** at **School of Mechanical and Manufacturing Engineering**, National University of Sciences and Technology, Islamabad.

Student Name: Muhammad Abdullah

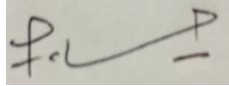
Signature: 

Examination Committee:

a) Internal Examiner 1: Dr. Umer Asgher
(Assistant Professor SINES)

Signature: 

b) Internal Examiner 2: Dr. Khawaja Fahad Iqbal
(Assistant Professor SMME)

Signature: 

Supervisor Name: Dr. Yasar Ayaz

Signature: 

Name of HOD: Dr. Kunwar Faraz Ahmed Khan

Signature: 

AUTHOR'S DECLARATION

I Muhammad Abdullah hereby state that my MS thesis titled “**Human Reliability Analysis using AI**” is my own work and has not been submitted previously by me for taking any degree from National University of Sciences and Technology, Islamabad or anywhere else in the country/world.

At any time if my statement is found to be incorrect even after I graduate, the university has the right to withdraw my MS degree.

Name of Student: Muhammad Abdullah

Date: 14-10-2024


PLAGIARISM UNDERTAKING

I solemnly declare that research work presented in the thesis titled “**Human Reliability Analysis using AI**” is solely my research work with no significant contribution from any other person. Small contribution/ help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero-tolerance policy of the HEC and National University of Sciences and Technology (NUST), Islamabad towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS degree, the University reserves the rights to withdraw/revoke my MS degree and that HEC and NUST, Islamabad has the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized thesis.

Student Signature: _____

A handwritten signature in blue ink, appearing to read 'Muhammad Abdullah', written over a horizontal line. The signature is stylized with a large loop at the beginning and a vertical stroke at the end.

Name: Muhammad Abdullah

DEDICATION

Dedicated to my exceptional parents and adored siblings whose tremendous support and cooperation led me to this wonderful accomplishment.

ACKNOWLEDGEMENTS

With the boundless grace and wisdom of Allah Subhanahu wa Ta'ala, I draw my strength and guidance. It is through His divine mercy that I have been blessed and guided throughout this academic journey and in every aspect of my life. To Him, I offer my deepest gratitude, for it is by His will that I have been fortunate to be surrounded by individuals who have been steadfast pillars of support and guidance.

I extend my heartfelt appreciation to my parents, who have been an endless source of love, resilience, and unwavering support. Their unconditional faith in me and their prayers have been the foundation upon which I have built my aspirations.

To my beloved siblings, who have stood by me through every challenge, your faith in me and your constant encouragement have been truly invaluable. I am also profoundly grateful to my supervisor, Dr. Yasar Ayaz, who has exemplified mentorship at its finest. His expertise, insights, and commitment have played a significant role in shaping this work. Throughout the course of this thesis, his invaluable guidance, constructive feedback, and remarkable patience have been instrumental in transforming my aspirations into reality.

I extend special thanks to my friends, who have been a constant source of comfort, joy, and reassurance. Their unwavering presence and belief in my abilities, particularly during moments of self-doubt, have been incredibly uplifting.

Lastly, to all who have contributed, even in the smallest way, to my academic journey, I offer my sincere gratitude. Whether it was a word of encouragement, a supportive gesture, or valuable advice, I hold it in high regard. To each and every one of you, may Allah bestow His countless blessings upon you, and may He guide us all in our endeavors.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	9
TABLE OF CONTENTS.....	10
LIST OF FIGURES	14
LIST OF TABLES	15
ABSTRACT.....	16
CHAPTER 1: INTRODUCTION.....	17
1.1. Fake News	17
1.1.1. Causes of Fake News.....	17
1.1.2. Motivation Behind the Spread of Fake News	18
1.1.3. Effects of Fake News.....	20
1.1.4. Impact of Fake News on Different Age Groups.....	21
1.1.5. Strategies for Combating Fake News	22
1.2. Background	24
1.3. Motivation	25
1.4. Challenges	26
1.4.1. Proliferation of Misinformation.....	26
1.4.2. Challenges in Classifying Fake News	26
1.4.3. Scarcity of Labeled Data	26
1.4.4. Bias in News Sources	27
1.4.5. Difficulty in Tracking Fake News	27
1.4.6. Need for Rapid and Precise Detection Methods.....	27
1.4.7. Language Use in Fake News	28
1.4.8. Mixing True Stories with False Details	28
1.4.9. Limited Fake News Data	28
1.5. Objectives	29
1.5.1. Advanced Preprocessing Techniques:	29
1.5.2. Utilization of Ensemble and Deep Learning Methods:	29
1.5.3. Developing a Resilient Detection Framework:	29

1.5.4. Performance Evaluation and Comparison:	29
1.6. Proposed Solution.....	30
1.7. Thesis Organization.....	31
CHAPTER 2: LITERATURE REVIEW	32
2.1. Early Instances of Fake News and Its Roots in Traditional Media.....	32
2.2. The Rise of Fake News with the Advent of Social Media (2010s).....	32
2.3. Early Attempts to Combat Fake News: Initial Research and Approaches (2017-2018)..	33
2.4. Innovations in Feature Integration and Model Complexity (2018-2019)	34
2.5. The Emergence of Transformer Models and Their Impact (2019-2020).....	34
2.5. The Evolution of Multi-Task Learning and Contextual Integration (2021-2024)	35
2.6. Current Challenges and Ongoing Research Directions.....	36
2.7. Future Directions and Potential Solutions.....	37
2.8. Literature Summary.....	38
2.4. Research Gap.....	38
CHAPTER 3: PROPOSED SYSTEM	40
3.1. Dataset.....	40
3.2. Deep Learning and Machine Learning Models.....	40
3.2.1. BERT (Bidirectional Encoder Representations from Transformers)	41
3.2.2. LSTM (Long Short-Term Memory Networks)	42
3.2.3. Bi-LSTM (Bidirectional LSTM)	43
3.2.4. Random Forest.....	44
3.2.5. Decision Trees	46
3.2.6. Artificial Neural Network (ANN)	47
3.2.7. Support Vector Machine (SVM)	48
3.2.7. Logistic Regression	49
3.3. Proposed System Diagram	50
CHAPTER 4: IMPLEMENTATION	52
4.1. Data Collection and Preprocessing	52
4.1.1. Overview of the LIAR Dataset.....	52
4.1.3. Dataset Statistics.....	54
4.1.4. Topic Distribution and Truthfulness.....	54

4.1.5. Challenges in Data Analysis.....	57
4.1.6. Label Definitions	57
4.2. Data Preprocessing Techniques	57
4.2.1. Tokenization and Stopword Removal	58
4.2.2. Stemming.....	58
4.2.3. Part-of-Speech (POS) Tagging.....	59
4.2.4. Clipping and Padding	60
4.2.5. Word Embeddings	60
4.2.6. Feature Extraction.....	61
4.2.7. N-grams	61
4.3. Methodology	62
4.3.1. Baseline Models	62
4.3.2. Proposed Model Architecture	63
1. The Hybrid Model:	63
2. Random Forest:	65
3. Bi-Directional Long Short-Term Memory (bi-LSTM) Network:.....	66
4. Decision Trees:.....	67
5. Artificial Neural Network (ANN):.....	67
6. Support Vector Machine (SVM):.....	68
7. Logistic Regression:.....	68
4.3.3. Training and Optimization.....	69
1. Data Preparation:	69
2. Optimization Techniques:.....	69
3. Performance Metrics:.....	70
CHAPTER 5: RESULTS AND DISCUSSION.....	71
5.1. Binary Classification Results	71
5.1.1. LSTM:	71
5.1.2. Random Forest:.....	72
5.1.3. Logistic Regression:	72
5.1.4. SVM:	73
5.2. Six-Way Classification Results.....	73

5.2.1. Bi-LSTM and LSTM:	73
5.2.2. SVM and ANN:	73
5.2.3. Random Forest and Decision Trees:	74
5.2.4. Logistic Regression:	75
5.2.5. The Hybrid Model:	76
5.2.6. Analysis of Confusion Matrix	76
5.3. Performance Comparison with Existing Models	77
5.3.1. LSTM-Attention:	77
5.3.2. MMFD:	78
5.3.3. Memory-Network:	78
5.3.4. FDML:	79
CHAPTER 6: CONCLUSION AND FUTURE WORK	80
6.1. Conclusion.....	80
6.2. Future Work	81
6.2.1. Enhancing the Model with Large Language Models (LLMs).....	81
6.2.2. Incorporating Multilingual Capabilities	81
6.2.3. Real-Time Fake News Detection.....	81
6.2.4. Handling Data Imbalance with Advanced Techniques	82
6.2.5. Improving Explainability and Interpretability	82
REFERENCES	83

LIST OF FIGURES

Figure 1 Architecture of a BERT model.....	41
Figure 2 Architecture of an LSTM model	43
Figure 3 Architecture of a Bi-LSTM model	44
Figure 4 Architecture of a random forest model.....	45
Figure 5 Architecture of a decision tree.....	46
Figure 6 Architecture of the ANN model	47
Figure 7 Schematic Diagram of SVM architecture.....	49
Figure 8 Architecture of the logistic regression model.....	50
Figure 9 System architecture of the proposed hybrid model	51
Figure 10 Distribution of Labels in LIAR Dataset	55
Figure 11 Distribution of statement lengths in the complete dataset.....	55
Figure 12 Distribution of statement lengths in terms of labels.....	56
Figure 13 Most frequent words in the dataset.....	56
Figure 14 System architecture of the proposed hybrid model	63
Figure 15 Flow diagram of the random forest for binary classification	65
Figure 16 Flow diagram of LSTM for six-way classification	66
Figure 17 Confusion matrix for random forest in binary classification.....	72
Figure 18 Confusion matrix of random forest for six-way classification.....	74
Figure 19 Confusion matrix of decision tree for six-way classification.....	75
Figure 20 Confusion matrix of logistic regression for six-way classification	75
Figure 21 Confusion matrix of the hybrid model for six-way classification.....	76

LIST OF TABLES

Table 1: LIAR data statistics.....	52
Table 2: Data analysis.....	53
Table 3: An example of the LIAR dataset	53
Table 4: Top 24 news topics	54
<i>Table 5: Models' performances</i>	<i>71</i>
<i>Table 6: Performance evaluation of fake news detection methods.....</i>	<i>78</i>

ABSTRACT

The exponential growth of social media platforms has revolutionized the way information is shared and consumed, leading to the widespread dissemination of both factual and false information. The rapid spread of misleading or entirely false news poses a significant threat to public discourse and the integrity of democratic processes. The task of accurately classifying the truthfulness of statements is complex, particularly due to the nuanced and often ambiguous nature of content shared on social media. Traditional Natural Language Processing (NLP) techniques, such as Bidirectional Encoder Representations from Transformers (BERT), have demonstrated proficiency in contextual understanding and text classification. However, these approaches frequently encounter limitations in accuracy, largely due to their difficulties in managing imbalanced datasets and the lack of integration with supplementary feature sets. To address these challenges, this research proposes a novel hybrid model that combines the strengths of BERT with dependency parsing and integrates a Deep Learning (DL) model designed to process metadata. This hybrid approach enhances the model's ability to accurately analyze and classify the complex and varied structures within the dataset, leading to improved overall accuracy. Additionally, this study explores different network architectures and preprocessing techniques aimed at optimizing the model's performance. The proposed hybrid model was tested on the LIAR dataset, achieving a notable 64.6% accuracy, which represents a 13.8% improvement over the previous leading method, the Fake News Detection Multi-Task Learning (FDML) model. The findings from this research indicate that the incorporation of richer linguistic features and metadata into classification models can significantly enhance the effectiveness of fake news detection and categorization on social media platforms.

Key Words: *Misleading news, syntactical nuances, BERT, binary classification, six-way classification.*

CHAPTER 1: INTRODUCTION

1.1. Fake News

Fake news refers to intentionally fabricated information or deceptive content designed to mislead readers or manipulate public perception [1]. The term has gained significant attention in recent years, particularly with the rise of social media and digital platforms where information can be disseminated rapidly and without stringent editorial oversight. Unlike misinformation, which can be the result of unintentional errors, fake news is deliberately created with the intent to deceive [1] [2].

1.1.1. Causes of Fake News

The causes of fake news are multifaceted and can be attributed to various social, political, economic, and technological factors.

- **Technological Advancements and Social Media:**

The internet and social media platforms have revolutionized the way information is produced, shared, and consumed. Unlike traditional media, where content typically undergoes editorial review, social media allows anyone to create and disseminate information, regardless of its accuracy [1]. The democratization of content creation has made it easier for individuals and organizations to spread fake news without the checks and balances that are usually present in traditional media outlets [2] [3].

- **Monetary Incentives:**

Fake news can be a lucrative business. Websites that generate sensationalist or misleading content often attract large amounts of web traffic, which can be monetized through advertising revenue [3]. Clickbait headlines and sensationalist stories are designed to attract attention, leading to more clicks, more ad impressions, and ultimately, more revenue for the creators. The profitability of fake news has led to the creation of "content farms" where misleading or entirely fabricated news stories are mass-produced to drive web traffic and generate income [2] [4] [5].

- **Political Propaganda:**

Fake news is often used as a tool for political propaganda. During elections or politically charged events, fake news can be strategically deployed to influence public opinion, undermine political opponents, or destabilize democratic processes. This form of fake news is often spread by political actors or organizations with the intent of swaying voters or manipulating political outcomes. Governments, political parties, or interest groups may use fake news to promote their agendas, create division among the electorate, or discredit opponents [6].

- **Psychological and Cognitive Factors:**

Human psychology plays a significant role in the spread of fake news. Cognitive biases, such as confirmation bias, where people tend to favor information that confirms their preexisting beliefs, make individuals more likely to believe and share fake news. Additionally, the emotional appeal of fake news—whether it elicits fear, anger, or surprise—can make it more likely to be shared. The sensational nature of fake news often triggers an emotional response, leading people to spread it without critically evaluating its accuracy [6].

- **Information Overload and Lack of Media Literacy:**

The sheer volume of information available on the internet can overwhelm individuals, making it difficult to discern credible sources from unreliable ones. This information overload, combined with a lack of media literacy, means that many people are ill-equipped to critically evaluate the content they encounter. The fast-paced nature of news consumption, especially on social media, encourages the rapid sharing of information, often without fact-checking or verifying the source [6] [7].

1.1.2. Motivation Behind the Spread of Fake News

The motivations for spreading fake news are as varied as its causes, ranging from personal gain to broader societal manipulation.

- **Financial Gain:**

As previously mentioned, the most direct motivation for spreading fake news is financial profit. Sensational and misleading headlines attract clicks, and clicks generate advertising revenue. This financial incentive can lead individuals or groups to create and spread fake news purely for profit, without regard for the social or ethical implications [1] [5].

- **Political Influence:**

Political motivations are a major driver of fake news. In politically polarized environments, fake news can be used to promote specific ideologies, discredit opponents, or sway public opinion. For example, during election campaigns, fake news stories may be created to undermine a candidate's reputation or to exaggerate the achievements of another. The strategic use of fake news in politics can manipulate electoral outcomes and erode trust in democratic institutions [2] [3].

- **Social and Cultural Influence:**

Fake news can also be spread to influence social and cultural narratives. For instance, fake news may be used to perpetuate stereotypes, fuel social tensions, or spread misinformation about certain groups. This can lead to increased polarization within society, as people become more entrenched in their beliefs and less willing to engage with opposing viewpoints [1] [3] [7].

- **Disruption and Destabilization:**

In some cases, the intent behind spreading fake news is to disrupt and destabilize society. This could be the goal of foreign actors seeking to undermine the stability of another country, or domestic groups aiming to create chaos and uncertainty. By spreading fake news, these actors can sow discord, weaken social cohesion, and challenge the integrity of democratic processes [5] [8].

- **Psychological Satisfaction:**

On an individual level, some people may spread fake news because it provides them with psychological satisfaction. Sharing sensational or alarming news can give individuals a sense of importance or influence within their social circles. Additionally, fake news that aligns with an individual's beliefs can provide cognitive reinforcement, making them feel validated and more likely to share the content [3] [5].

1.1.3. Effects of Fake News

The effects of fake news are widespread and can have serious implications for individuals, communities, and society as a whole.

- **Erosion of Trust in Media and Institutions:**

One of the most significant effects of fake news is the erosion of trust in media and institutions. When people are repeatedly exposed to fake news, it can lead to scepticism about the reliability of all news sources, including credible ones [8]. This distrust extends beyond the media to other institutions, such as government, academia, and science, undermining public confidence in these pillars of society [5] [8].

- **Polarization and Division:**

Fake news contributes to social and political polarization by reinforcing existing biases and creating echo chambers where people are only exposed to information that aligns with their beliefs. This can deepen divisions within society, as individuals become more entrenched in their views and less open to dialogue with those who hold opposing perspectives. The spread of fake news can create an "us versus them" mentality, where different groups within society are pitted against each other [9].

- **Influence on Elections and Democratic Processes:**

The impact of fake news on elections and democratic processes is a growing concern. Fake news can influence voter behaviour, sway election outcomes, and undermine the integrity of the democratic process. For example, during the 2016 U.S. Presidential election, fake news stories were widely circulated, with some analysts suggesting that they may have influenced the outcome of the election [10]. The use of fake news to manipulate public opinion poses a significant threat to the fairness and transparency of democratic systems.

- **Public Safety and Health Risks:**

Fake news can also have serious consequences for public safety and health. During the COVID-19 pandemic, for example, misinformation about the virus, its origins, and potential treatments spread rapidly on social media. This led to confusion, panic, and in some cases, harmful behaviours, such as people taking unproven and dangerous remedies. Fake news related to health and safety can lead to widespread fear and panic, potentially endangering lives [11] [12].

- **Economic Impact:**

The spread of fake news can have economic repercussions, particularly for businesses and financial markets. Companies targeted by fake news may suffer damage to their reputation, leading to loss of customers and revenue. In financial markets, fake news can cause stock prices to fluctuate, leading to volatility and potentially significant financial losses for investors [2] [13].

- **Impact on Mental Health:**

The constant exposure to fake news, especially sensational and alarming stories, can have a negative impact on mental health. Individuals may experience increased anxiety, stress, and a sense of helplessness when faced with a barrage of misleading or false information. The emotional toll of fake news can contribute to a sense of disillusionment and disconnection from society [14].

1.1.4. Impact of Fake News on Different Age Groups

Fake news affects people of all ages, but its impact can vary depending on the demographic.

- **Children and Adolescents:**

Young people, particularly those in their formative years, are highly impressionable and may not yet have developed the critical thinking skills necessary to discern fake news from credible information. Exposure to fake news can shape their beliefs and perceptions in ways that are difficult to undo [15]. For example, children who are exposed to fake news related to health or science may develop misconceptions that persist into adulthood. Additionally, the spread of fake news among young people can contribute to bullying, social exclusion, and the reinforcement of harmful stereotypes [6] [16].

- **Adults:**

Adults are also susceptible to fake news, particularly when it aligns with their preexisting beliefs or biases. This demographic may be more likely to share fake news within their social networks, contributing to its spread [16]. The impact of fake news on adults can be seen in the political sphere, where it can influence voting behaviour and opinions on social issues. Additionally, adults who are constantly exposed to fake news may experience heightened stress and anxiety, particularly if the news relates to health, safety, or financial matters [6].

- **Older Adults:**

Older adults are often targeted by fake news due to a combination of factors, including lower levels of digital literacy and a tendency to trust information that appears in traditional media formats. This demographic is particularly vulnerable to scams and misinformation, which can have severe consequences, such as financial loss or damage to their health. The spread of fake news among older adults can also contribute to isolation and distrust, as they may become sceptical of all information sources [6] [16].

1.1.5. Strategies for Combating Fake News

Given the serious implications of fake news, it is essential to develop strategies to combat its spread and mitigate its impact.

- **Improving Media Literacy:**

One of the most effective ways to combat fake news is by improving media literacy among the public. This involves educating individuals on how to critically evaluate the information they encounter, identify credible sources, and understand the mechanisms behind news production and distribution [16]. Educational programs that focus on media literacy can be integrated into school curricula and community workshops, targeting all age groups. By equipping people with the skills to discern fact from fiction, society can build resilience against the spread of misinformation [16] [17].

- **Fact-Checking Initiatives:**

Fact-checking organizations play a crucial role in debunking fake news. These entities systematically analyze questionable content, provide evidence-based corrections, and publish their findings. Collaborations between social media platforms and fact-checking organizations can help flag and reduce the visibility of fake news [17]. Fact-checking tools, such as browser extensions or integrated features within social media apps, can also empower users to verify the credibility of information in real time [18].

- **Technology and AI Solutions:**

Advances in artificial intelligence (AI) and machine learning have enabled the development of tools designed to detect and filter out fake news. Algorithms can be trained to recognize patterns commonly found in fake news, such as sensationalist language, lack of credible sources, and abnormal sharing patterns [16]. Social media platforms can use AI

to automatically flag or remove content identified as fake news. Additionally, blockchain technology could be explored as a means to ensure the authenticity of news sources, creating a transparent and tamper-proof system for verifying information [16] [18].

- **Regulatory Measures:**

Governments and regulatory bodies can introduce legislation aimed at curbing the spread of fake news. This might include laws that hold platforms accountable for the content they host or penalties for individuals and organizations that deliberately create and disseminate false information. However, regulatory measures must balance the need to combat fake news with the protection of free speech to avoid censorship and the suppression of legitimate discourse [6] [16].

- **Community Engagement and Awareness Campaigns:**

Raising public awareness about the dangers of fake news is essential. Community-driven initiatives, such as public service announcements, social media campaigns, and workshops, can educate the public about the impact of fake news and encourage responsible sharing of information. By fostering a culture of skepticism and critical thinking, communities can become more vigilant in identifying and combating fake news [16] [19].

- **Collaboration with Influencers and Opinion Leaders:**

Influencers and opinion leaders have significant reach and can play a role in countering fake news. By promoting accurate information and debunking myths, these individuals can help shift public perception and reduce the spread of misinformation. Partnerships with credible figures in various fields, including science, health, and politics, can amplify efforts to combat fake news [16] [19].

To summarize the whole discussion; Fake news presents a complex and multifaceted challenge that has far-reaching implications for society. Its causes are rooted in technological advancements, economic incentives, political agendas, psychological factors, and a general lack of media literacy [1] [2] [3] [4]. The motivations for spreading fake news range from financial gain and political influence to social disruption and psychological satisfaction [5] [6] [11]. The effects of fake news are profound, eroding trust in media and institutions, deepening social and political divisions, influencing elections, and posing risks to public safety and health [5] [8] [9] [12].

Different age groups are impacted by fake news in varying ways, with children and adolescents being particularly vulnerable due to their developing cognitive abilities, while older adults may fall prey to scams and misinformation due to lower digital literacy [6] [10] [11] [12]. The societal damage caused by fake news underscores the urgent need for comprehensive strategies to combat its spread [3] [7] [13] [16].

Addressing the issue of fake news requires a multi-pronged approach that includes improving media literacy, implementing fact-checking initiatives, leveraging technology, enacting regulatory measures, and engaging communities [6] [15] [16]. By fostering a more informed and discerning public, society can mitigate the harmful effects of fake news and protect the integrity of information in the digital age [6] [14] [16] [17].

The fight against fake news is ongoing, and it requires the collective efforts of individuals, communities, governments, and technology companies [18] [19]. As information continues to flow at unprecedented speeds, the ability to discern truth from falsehood will be critical in maintaining a well-informed and cohesive society [4] [16] [18].

1.2. Background

The rapid proliferation of social media has fundamentally transformed how information is shared and consumed on a global scale. Platforms like Facebook, Twitter, and Instagram have empowered users to disseminate news and opinions with unprecedented speed and reach [20] [21]. While this has democratized information access, it has also facilitated the widespread dissemination of misinformation, commonly known as "fake news." Fake news refers to deliberately false information intended to mislead readers, and it poses significant challenges to public trust and societal stability [10]. This issue becomes particularly critical during major events, such as elections or pandemics, where the spread of false information can have severe real-world consequences [22].

One of the major hurdles in combating fake news is the accurate classification of statements based on their truthfulness. Categories such as True, Mostly True, Half True, Barely True, False, and Completely False (often referred to as "Pants on Fire") are used to differentiate the varying degrees of truth [7]. However, this task is complicated by the skewed nature of the data, with true statements often outnumbered by various levels of falsehoods, largely due to the prevalence of fake news on social media [9]. Traditional methods, like manual fact-checking by journalists, are

insufficient to manage the vast volume and speed of information flow on these platforms. Thus, automating this process is crucial for efficiently identifying and mitigating the impact of false information [4] [11] [23].

Recent advancements have seen the development of machine learning and deep learning techniques aimed at automating fake news detection. For instance, Wang et al. introduced a convolutional neural network (CNN) model that achieved 27.4% accuracy on the LIAR dataset, which comprises short statements labelled with truthfulness ratings [24]. Building on this, Kirilin et al. improved accuracy to 45.7% using the Speaker2Credit method, which incorporated additional features into a long short-term memory (LSTM) model [25]. Similarly, Long et al. applied attention mechanisms within LSTM models, achieving 41.5% accuracy [26]. Other researchers have explored transformer models, such as BERT, to capture more contextual information and enhance classification performance [27].

Despite these advancements, current models face several limitations. They often struggle with imbalanced and diverse datasets and may fail to generalize well across different contexts and topics [28] [29]. Additionally, essential syntactical features, such as dependency parsing, constituency parsing, and part-of-speech (POS) tagging, are frequently overlooked, even though they are crucial for understanding the nuances of human language [11] [17] [22]. While the integration of richer linguistic features has been proposed, it has not been fully exploited. Furthermore, these models do not adequately address the dynamic nature of language and the evolving patterns of misinformation, limiting their long-term effectiveness [8] [26] [30].

1.3. Motivation

The detection of fake news is critical in today's digital landscape, where misinformation spreads rapidly and widely, influencing public opinion and decision-making [1] [6] [16]. Fake news undermines trust in credible information sources, fuels social and political polarization, and can lead to real-world consequences such as public health risks, financial losses, and compromised democratic processes [5] [22] [23].

Effective fake news detection is essential to safeguard the integrity of information, ensure an informed public, and maintain the credibility of legitimate news outlets. As digital platforms continue to evolve, the need for robust, rapid, and precise detection methods grows, making this an urgent area for research and technological innovation. Addressing this challenge not only

protects individuals from deception but also upholds the overall health of our information ecosystem.

1.4. Challenges

There are several challenges regarding the mitigation of fake news which are explained below:

1.4.1. Proliferation of Misinformation

The spread of misinformation has become a significant concern, especially with the rise of digital media platforms that allow information to be shared rapidly and widely. This proliferation makes it challenging to control false narratives and can have serious consequences, including public mistrust, social division, and the spread of harmful beliefs [1] [13] [31]. As fake news continues to circulate, it becomes harder for individuals to distinguish between credible and misleading information, undermining the integrity of reliable sources and creating confusion on critical issues [10] [16] [25].

1.4.2. Challenges in Classifying Fake News

Classifying fake news is inherently difficult due to its complex and evolving nature. Fake news is not a single, uniform entity; it can range from outright falsehoods to misleading or partially true information, as well as satire that is not meant to deceive but can still misinform. Rashkin's analysis points out that fake news often uses subjective, intensifying, or hedging language to dramatize or obscure facts, which complicates classification [31] [32]. Traditional methods that rely on feature-based approaches struggle because the linguistic nuances of fake news can vary widely, requiring sophisticated and often labour-intensive techniques to identify accurately [3] [32] [33].

1.4.3. Scarcity of Labeled Data

One of the key challenges in addressing fake news is the scarcity of labelled data, which is essential for training machine learning models to detect misinformation. Currently, most available datasets focus on political fake news, leaving other areas, such as health, science, or finance, largely unexplored [6] [16] [25]. This lack of diverse data hampers the development of robust detection

systems that can operate effectively across different domains. Without sufficient labelled examples of fake news in various contexts, it's challenging to develop comprehensive models that generalize well beyond the political sphere [16] [34].

1.4.4. Bias in News Sources

Bias in news sources is another significant issue, as it can skew the perception of events and facts, contributing to the spread of misinformation. Fake news often exploits these biases, amplifying polarized views to mislead specific audiences [4] [16] [31]. This makes it difficult not only to identify fake news but also to disentangle it from biased yet factually accurate reporting. When algorithms trained on biased data attempt to classify news, they may inherit these biases, leading to inaccurate or skewed results that fail to address the root problem [35] [36].

1.4.5. Difficulty in Tracking Fake News

Tracking fake news as it spreads across different platforms is a formidable challenge due to the rapid and often viral nature of digital content. Fake news can be shared, reposted, and altered as it moves from one medium to another, making it difficult to trace back to the original source or measure its impact accurately [14] [19]. The dynamic nature of fake news, with content constantly evolving and adapting, requires sophisticated tracking tools that can monitor changes and dissemination patterns in real time [14] [33] [37].

1.4.6. Need for Rapid and Precise Detection Methods

Given the speed at which fake news spreads, there is a critical need for detection methods that are both rapid and precise. Delays in identifying fake news can allow it to reach large audiences, causing significant damage before corrections or clarifications can be issued. Current detection systems often struggle with the dual demands of speed and accuracy, particularly when dealing with subtle or partially true misinformation [33] [37]. The ideal solution would involve real-time analysis capable of catching and flagging fake news immediately, but this remains a significant technical challenge, requiring ongoing research and development to achieve reliable results [4] [13] [16].

1.4.7. Language Use in Fake News

Rashkin's research highlights how language plays a key role in differentiating fake news from true news across various forms, such as satire, hoax, and propaganda. Fake news often utilizes subjective, intensifying, or hedging language to obscure facts, dramatize events, or sensationalize stories, which complicates the task of distinguishing it from legitimate news [18]. The deliberate use of vague or emotionally charged language is intended to mislead or confuse readers, making it more challenging for automated systems to accurately classify content as fake or true [5] [31] [38].

1.4.8. Mixing True Stories with False Details

A common tactic in fake news is to mix true stories with false details, creating content that is difficult to recognize as entirely false. This blend of truth and fiction, such as including accurate numerical data alongside misleading contextual information, can make the fabricated parts less noticeable [15] [31]. For example, citing a correct figure but attributing it to an incorrect source or event can create a deceptive narrative that appears credible at first glance [6]. This strategy exploits the trust readers place in factual elements, drawing attention to the accurate details while the falsehoods go unnoticed [15] [17] [24].

1.4.9. Limited Fake News Data

The current limitation in fake news data, particularly outside the political domain, poses a challenge for developing comprehensive detection methods. Most available datasets are focused on political misinformation, leaving a gap in other areas like health, science, and finance [15]. This lack of diverse, labelled data restricts the ability of detection systems to generalize and adapt to different types of fake news, highlighting a significant area for future research and development. Expanding the scope of data collection to include a broader range of topics is essential for building more versatile and effective detection tools [17] [20] [22].

1.5. Objectives

1.5.1. Advanced Preprocessing Techniques:

Enhance the quality of text data by implementing advanced preprocessing methods that improve the understanding of syntactical structures and contextual meanings. This includes techniques like stemming, lemmatization, and the removal of noise, which help refine the input data for better analysis and classification.

1.5.2. Utilization of Ensemble and Deep Learning Methods:

Employ ensemble methods, such as Random Forest, and advanced deep learning models like bidirectional Long Short-Term Memory (bi-LSTM). These approaches help manage imbalanced datasets by combining multiple algorithms to improve accuracy and capture the complex patterns inherent in fake news, leading to more reliable detection.

1.5.3. Developing a Resilient Detection Framework:

Build a robust framework for fake news detection by integrating metadata features (e.g., publication date, source, and author information) alongside textual data. Employing effective extraction techniques ensures that additional contextual information is utilized, enhancing the overall detection capability.

1.5.4. Performance Evaluation and Comparison:

Systematically evaluate and compare various machine learning and deep learning algorithms to identify the most effective approach for fake news classification. This involves testing models on key performance metrics like accuracy, precision, recall, and F1-score, to ensure the chosen methods meet the required standards for real-world application.

1.6. Proposed Solution

To address these challenges, we propose an enhanced approach that improves upon the BERT model by incorporating POS tags, dependency parsing, and constituency parsing to capture more syntactical information [39] [40]. In our preprocessing pipeline, each statement is cleansed by removing stopwords and padded to a uniform length of 45 words, based on the maximum sentence length, excluding outliers. The model leverages GloVe embeddings for word representation, POS tags for syntactical features, and a deep neural network architecture for classification [39] [41]. This comprehensive approach enables the model to better understand the structure and meaning of statements, resulting in substantial improvements in classification accuracy and robustness, particularly when dealing with complex and nuanced data.

Our work aims to significantly advance the field of fake news detection by overcoming the limitations of existing approaches in capturing syntactical features and by introducing richer linguistic elements. This development offers a more robust tool for mitigating the spread of misinformation on social media platforms.

In this research, our contributions are as follows:

1. **Enhanced Hybrid Model:** We propose an advanced hybrid model that integrates BERT with POS tags and dependency parsing, and a DL model for non-textual features to enhance fake news detection.
2. **Feature and Preprocessing Analysis:** We provide an in-depth analysis of the impact of various features and preprocessing techniques on model performance.
3. **Empirical Validation:** Through extensive experiments on the LIAR dataset, we demonstrate the effectiveness of our approach, achieving a significant improvement in accuracy, with our model reaching 64.6%, surpassing existing methods by 13.8%.

1.7. Thesis Organization

The thesis is organized as follows:

- **Chapter 2: Literature Review**

Provides an overview of existing research on fake news detection, emphasizing the strengths and weaknesses of current methods.

- **Chapter 3: Proposed Approach**

Outlines the design and development of the detection framework, including data preprocessing, model selection, and overall system architecture.

- **Chapter 4: Implementation**

Details the implementation process of the proposed approach, including the tools, technologies, and challenges faced during development.

- **Chapter 5: Results and Discussion**

Presents the evaluation results of the detection models, analysing their performance and effectiveness in identifying fake news across various datasets.

- **Chapter 6: Conclusion and Future Work**

Summarizes the key findings of the study, discusses the implications for combating misinformation, and suggests potential directions for future research.

CHAPTER 2: LITERATURE REVIEW

2.1. Early Instances of Fake News and Its Roots in Traditional Media

The concept of fake news is not a recent phenomenon, though its impact and reach have been significantly amplified in the digital age. Historically, fake news has its roots in sensational journalism and propaganda disseminated through traditional media [2] [9]. These early forms of fake news were limited in scope and impact due to the constraints of print and broadcast media, which had inherent gatekeeping mechanisms. Editors and producers served as filters, ensuring that only content deemed fit for publication or broadcasting made it to the public [1] [4] [6].

In the past, fake news was often tied to the political agendas of those in power, who used propaganda to manipulate public opinion and control narratives [7]. For example, during wartime, governments frequently used propaganda to bolster support for the war effort, sometimes by spreading exaggerated or false information [10]. However, the impact of such fake news was relatively contained due to the limited reach of the media outlets at the time. The situation began to change with the advent of the internet, which democratized content creation and distribution, removing many of the traditional barriers to entry for information dissemination [2] [6] [16] [22].

2.2. The Rise of Fake News with the Advent of Social Media (2010s)

The term "fake news" began to gain prominence in the early 2010s, largely coinciding with the rise of social media platforms such as Facebook, Twitter, and Instagram [11] [15]. Unlike traditional media, social media platforms provided an open arena where anyone could publish content without editorial oversight. This lack of regulation, combined with the virality of content on these platforms, created an environment ripe for the spread of misinformation [2] [7] [25].

The 2016 U.S. Presidential election marked a pivotal moment in the history of fake news. During this period, fake news stories proliferated on social media, often outperforming real news in terms of engagement and reach [11] [15] [24]. A study conducted by the Pew Research Center found that 64% of adults believed that fake news had caused significant confusion about basic facts of current events during the election period [9]. Researchers like Silverman (2016) highlighted how

fake news could spread rapidly on social media, sometimes even surpassing legitimate news in terms of visibility and engagement [18] [27].

This period also saw the emergence of sophisticated misinformation campaigns, often orchestrated by foreign actors or political interest groups [15] [36]. These campaigns were designed to influence public opinion, undermine democratic processes, and create division within society [8]. The sheer scale of these operations and their impact on the democratic process prompted a wave of research focused on understanding the mechanisms behind the spread of fake news and developing methods to combat it [6] [16] [38].

2.3. Early Attempts to Combat Fake News: Initial Research and Approaches (2017-2018)

As the problem of fake news became more apparent, researchers began to develop automated methods for detecting and combating it. Wang (2017) made a significant contribution to the field by introducing the LIAR dataset, a benchmark dataset specifically designed for fake news detection [7]. The LIAR dataset contains 12,836 manually labelled short statements, each categorized based on its truthfulness. This dataset became a foundational resource for the development of various machine learning models aimed at detecting fake news [7] [16] [33].

One of the early models developed using the LIAR dataset was a convolutional neural network (CNN) model, which achieved an accuracy of 27.4% [7]. This model laid the groundwork for subsequent research by demonstrating the potential of deep learning techniques in the field of fake news detection. However, the CNN model also highlighted the challenges associated with imbalanced data and the need for more sophisticated approaches to capture the nuances of human language [7] [33].

In 2017, Ruchansky et al. developed a hybrid model that combined recurrent neural networks (RNNs) with user behaviour analysis to improve the accuracy of fake news detection. This model achieved high accuracy rates on platforms like Twitter and Weibo, demonstrating the importance of considering user behaviour and social context in fake news detection. However, the model struggled to generalize across different social media platforms due to variations in user behaviour, indicating the need for models that could adapt to different contexts [14].

2.4. Innovations in Feature Integration and Model Complexity (2018-2019)

As research progressed, scholars began to explore more complex models that could better capture the nuances of language and context. Kirilin et al. (2018) introduced the Speaker2Credit model, which integrated speaker profiles to improve the accuracy of fake news detection [19]. By considering the credibility of the speaker alongside the content of the statement, this model aimed to provide a more comprehensive analysis of the likelihood that a statement was true or false. However, the model faced challenges in effectively utilizing complex metadata, highlighting the difficulty of balancing model complexity with interpretability [19].

Long et al. (2018) took a different approach by incorporating attention mechanisms into long short-term memory (LSTM) models [16]. Attention mechanisms allow models to focus on specific parts of a statement that are most relevant to determining its truthfulness, thereby improving performance. However, while the LSTM with attention mechanisms achieved an accuracy of 41.5%, it sometimes struggled to handle nuanced linguistic features, particularly those related to syntactical structure and context [16].

These studies laid the foundation for more sophisticated approaches to fake news detection, but they also revealed the limitations of early models, particularly in terms of their ability to handle diverse datasets and adapt to evolving fake news tactics [16] [22]. The need for models that could capture deeper linguistic features and generalize across different contexts became increasingly apparent [16] [41].

2.5. The Emergence of Transformer Models and Their Impact (2019-2020)

The introduction of transformer-based models marked a significant advancement in the field of natural language processing (NLP), including fake news detection [41]. Transformer models, such as Bidirectional Encoder Representations from Transformers (BERT), became popular for their ability to capture contextual information more effectively than previous models [16] [24] [41].

Devlin et al. (2018) demonstrated that BERT could outperform earlier models on various NLP tasks, including fake news detection [25]. BERT's ability to understand the context of a word within a sentence, rather than treating each word in isolation, made it particularly well-suited for detecting the subtleties of fake news. However, BERT also had its limitations, particularly when it came to understanding very long texts due to its input length constraints. This limitation meant

that while BERT could excel in many areas, it sometimes struggled with more complex or lengthy content, which is common in fake news [15] [25] [33].

In response to these challenges, Liu et al. (2019) introduced RoBERTa, an optimized version of BERT that improved accuracy in fake news classification [22]. RoBERTa made several modifications to the original BERT architecture, including longer training times and larger batch sizes, which allowed it to better capture the complexities of language. However, the increased computational resources required by RoBERTa made it less accessible for real-time applications on a large scale, limiting its practicality in some contexts [4] [6] [42].

2.5. The Evolution of Multi-Task Learning and Contextual Integration (2021-2024)

As the limitations of existing models became clearer, researchers began to explore multi-task learning and the integration of richer contextual information to enhance the accuracy and robustness of fake news detection models. Liao et al. (2021) proposed the Fake News Detection Multi-Task Learning (FDML) model, which combined topic labels with contextual information to improve performance. The FDML model's unique integration of representation learning and multi-task learning allowed it to capture both the content and context of news stories more effectively than previous models [33].

Building on the advancements made with models like FDML, recent research efforts have continued to push the boundaries of fake news detection by incorporating more sophisticated syntactical and contextual features [15] [26]. Fagundes et al. (2024) emphasized the importance of integrating syntactical features, such as context-free grammars and syntactic dependency trees, to improve detection accuracy [29]. Their systematic review of fake news detection methods highlighted how deeper syntactic representations can enhance the ability to understand and classify complex news content. By incorporating these advanced syntactical features, their approach sought to overcome some of the limitations associated with previous models, which often relied heavily on surface-level lexical features [29].

However, Fagundes et al.'s methods, while effective, also introduced new challenges, particularly in terms of computational intensity [6] [16]. The complex algorithms required to process and integrate these syntactical features demand significant computational resources, which can limit the scalability of these models in real-time applications. Moreover, these methods may not

generalize well across different languages and dialects, which presents a barrier to their widespread adoption [17] [42] [43].

In parallel, Agarwal et al. (2024) proposed a novel framework called Granularity-based Fake News Detection (GRAFED) [33]. This framework enhances fake news detection by combining both fine-grained word-level features and coarse-grained sentence-level features. GRAFED utilizes a combination of word vector representations, psycho-linguistic features, and sentiment analysis, using tools like the Empath library to analyze the emotional tone and psychological states conveyed in the text. The goal of GRAFED is to provide a more nuanced analysis of news content by considering both the micro-level details (such as specific word choices) and the macro-level context (such as the overall sentiment and psychological undertones of the content) [33].

Despite the innovative nature of GRAFED, its reliance on multiple feature sets can complicate the model training process [6] [33] [41]. The integration of diverse features requires careful tuning and extensive computational resources, which can make the model more challenging to implement across different datasets and platforms. Additionally, the complexity of GRAFED may hinder its ability to be adapted quickly to new forms of fake news, which are constantly evolving in response to detection methods [16] [33].

2.6. Current Challenges and Ongoing Research Directions

Even with the advancements in fake news detection technologies, several challenges remain that continue to hinder the effectiveness and generalization of existing models. One of the most significant issues is the diversity and imbalance within datasets. True statements are often outnumbered by false ones, making it difficult to maintain high accuracy across all categories of truthfulness. This imbalance can lead to models that perform well in detecting certain types of fake news while failing in others, reducing their overall reliability [41] [43] [44].

Moreover, the evolving nature of language and the continuous adaptation of fake news creators to bypass detection methods pose ongoing challenges. Fake news is not static; it changes in response to political, social, and technological shifts [41] [45]. For example, new forms of misinformation can emerge in response to current events, such as health crises or political upheavals, requiring models to adapt quickly to new patterns and linguistic nuances. This dynamic nature of fake news makes it difficult to develop models that can remain effective over time without frequent retraining and updating [38] [40] [42].

Another significant challenge is the integration of metadata features, such as the credibility of the news source, the history of the speaker, and the context in which the news is shared [37] [39]. While some models have begun to incorporate these features, effectively integrating them into the detection process remains complex [42]. Metadata can provide valuable context that enhances the accuracy of fake news detection, but it also introduces additional layers of complexity that must be carefully managed [43] [45].

2.7. Future Directions and Potential Solutions

Given the current state of research, several directions for future work have been identified that could further enhance the effectiveness of fake news detection models. One promising approach is the development of models that are better at handling imbalanced datasets. Techniques such as data augmentation, where new data is synthetically generated to balance the dataset, or the use of generative adversarial networks (GANs) to create realistic fake news examples, could help to mitigate the effects of data imbalance [37] [43] [46].

Another important direction is the incorporation of more robust linguistic and syntactical features. While significant progress has been made in this area, there is still room for improvement, particularly in the development of models that can capture the deeper nuances of human language. This could involve the use of more advanced natural language processing techniques, such as transformers that are specifically designed to handle long-range dependencies and complex sentence structures [31] [38] [41].

Furthermore, the integration of multi-modal data, such as images and videos, into fake news detection models could provide a more comprehensive approach to identifying misinformation. Many fake news stories are accompanied by misleading images or videos that reinforce the false narrative. By incorporating visual data into the detection process, models could gain a more holistic understanding of the content, improving their ability to detect fake news [16] [24] [39].

Finally, the development of real-time fake news detection systems remains a critical area of research. As fake news spreads rapidly across social media platforms, the ability to detect and respond to misinformation in real time is essential for mitigating its impact. This will require the development of more efficient algorithms that can process large volumes of data quickly without sacrificing accuracy [13] [25] [37].

2.8. Literature Summary

The field of fake news detection has evolved significantly over the past decade, with researchers making substantial progress in developing models that can automatically detect and classify fake news. From the early days of manual verification and simple machine learning models to the current state-of-the-art approaches that leverage deep learning and advanced natural language processing techniques, the journey has been marked by both successes and challenges.

Despite the advancements, current models are still limited by issues such as data imbalance, the evolving nature of language, and the complexity of integrating diverse features. As researchers continue to refine these models, it is likely that future innovations will focus on addressing these challenges through the development of more sophisticated algorithms, the incorporation of multi-modal data, and the creation of real-time detection systems.

Research in fake news detection is crucial for reducing misinformation and improving the reliability of social media content. As fake news threatens public trust and stability, developing effective detection methods is essential. Continued innovation and collaboration will drive new tools and strategies to combat misinformation.

2.4. Research Gap

In this research, we address several key gaps identified through our literature review, which are crucial for enhancing the effectiveness of fake news detection models:

1. Handling Data Imbalance:

- **Gap Identified:** Many existing models struggle with the skewed distribution of true versus false statements, leading to biased and inaccurate classification [7].
- **Our Contribution:** We address this issue by implementing advanced data handling techniques, including the use of ensemble methods like Random Forest and deep learning approaches such as bi-LSTM, to improve the model's robustness and accuracy across all truthfulness categories.

2. Incorporation of Syntactical Features:

- **Gap Identified:** Traditional models often overlook important syntactical features like dependency parsing, POS tagging, and constituency parsing, which are crucial for understanding the nuanced structure of language [25].

- **Our Contribution:** We enhance our model by integrating these richer linguistic features, allowing for a more detailed and accurate analysis of statement structure and meaning, which is particularly useful in distinguishing subtle differences between truthfulness levels.

3. Contextual Understanding and Metadata Integration:

- **Gap Identified:** Existing models tend to focus heavily on textual data while neglecting contextual information and metadata, such as speaker credibility and the context in which a statement is made [33].
- **Our Contribution:** We incorporate metadata and contextual features into our model, improving its ability to accurately classify statements by considering the broader context in which the information is presented, thus enhancing the overall detection capability.

By addressing these research gaps, our work not only improves the accuracy and reliability of fake news detection models but also contributes to the development of more adaptable, context-aware, and scalable solutions for mitigating the spread of misinformation in the digital age.

CHAPTER 3: PROPOSED SYSTEM

In this chapter, we will discuss the proposed system for detecting fake news. This includes an overview of the data collection process, and a detailed examination of the various deep learning model architectures employed in the detection framework.

3.1. Dataset

The LIAR dataset is a comprehensive collection of fact-checked news statements, widely used in fake news detection research [7]. It contains 12,836 short news statements, each of which has been manually labeled based on its truthfulness [7]. The dataset is sourced from **PolitiFact.com**, a renowned fact-checking website, where professional editors evaluate the accuracy of political statements and news, ensuring the reliability and precision of the labels. The dataset serves as a valuable resource for machine learning models, offering a wide range of truthfulness categories, metadata, and additional contextual information [7] [33] [34].

Each news statement in the LIAR dataset averages 17.9 tokens, making them short but information rich. These statements come from various sources, including political debates, news releases, and social media posts. This diversity in sources introduces challenges in handling different formats of news presentation but provides a comprehensive overview of information disseminated across various platforms. The news statements are attributed to 3,318 public speakers, providing an additional layer of metadata on the credibility and historical accuracy of the speakers [7] [33].

3.2. Deep Learning and Machine Learning Models

Deep learning models have significantly advanced text analysis and fake news detection, offering precise and efficient ways to understand complex linguistic patterns. In this research, a variety of machine learning and deep learning models were employed to detect fake news from the LIAR dataset. Each model brings its own strengths and limitations, and the selection of models aims to strike a balance between accuracy, complexity, and computational efficiency. Below is a detailed explanation of each model, including their technical workings, advantages, and disadvantages.

3.2.1. BERT (Bidirectional Encoder Representations from Transformers)

BERT is a state-of-the-art transformer-based model developed by Google, designed to handle NLP tasks by learning contextual relations between words in a text. Unlike traditional models that read text either left-to-right or right-to-left, BERT reads both directions simultaneously, thus being bidirectional. BERT can be fine-tuned for specific tasks like classification, including fake news detection, by utilizing its pre-trained weights on vast corpora of text [17] [23] [31] [46].

BERT uses transformers, which are attention mechanisms that allow the model to focus on different parts of a sentence while processing it. It achieves this through multiple layers of encoders that compute self-attention scores between tokens (words) in a sentence. BERT processes sentences in parallel (unlike RNNs), making it efficient for large-scale text processing.

- **Input:** Tokenized text, padded to ensure consistency in input dimensions.
- **Architecture:** Multiple layers of transformer encoders.
- **Output:** Encoded representations of words in a context-aware fashion.

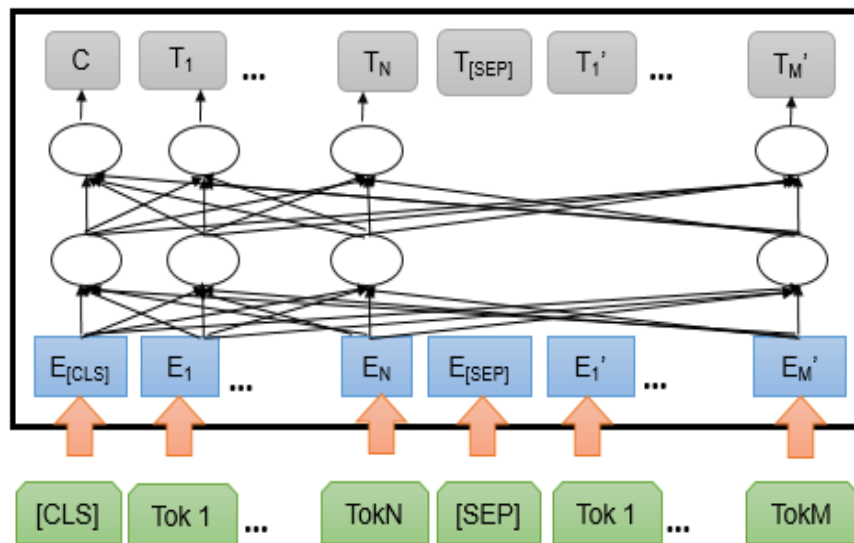


Figure 1 Architecture of a BERT model

Pros:

- **Contextual Understanding:** BERT excels at understanding the meaning of words within the context of a sentence, making it ideal for tasks that require nuanced comprehension.

- **Transfer Learning:** It can be fine-tuned for a variety of NLP tasks, making it versatile.
- **Bidirectional Attention:** Allows the model to capture context from both directions in a sentence, improving performance in understanding complex sentence structures.

Cons:

- **Computationally Expensive:** Training and fine-tuning BERT requires significant computational resources.
- **Limited by Input Length:** BERT is constrained by a maximum token length (typically 512 tokens), which can limit its ability to handle very long texts.

3.2.2. LSTM (Long Short-Term Memory Networks)

LSTM is a type of recurrent neural network (RNN) specifically designed to learn long-term dependencies in sequential data. It is highly suitable for tasks like text classification where the order of words is important. In this research, LSTM is employed to capture relationships in text data over time [4] [16] [44] [45].

LSTM has a unique cell structure that consists of gates—input, forget, and output gates—that regulate the flow of information through the network. This allows the network to "remember" relevant information and "forget" irrelevant data across long sequences, which addresses the vanishing gradient problem that affects traditional RNNs [16] [44].

- **Input:** Sequential data (in this case, tokenized text).
- **Architecture:** LSTM cells with memory blocks and gates.
- **Output:** Contextual representations of each word in a sequence, considering the past information.

Pros:

- **Effective for Sequential Data:** LSTMs are highly effective for text data where understanding word order and context over long sequences is important.
- **Captures Long-Term Dependencies:** The memory cells allow the model to maintain information over longer periods.

Cons:

- **Slow Training:** Due to the sequential nature of LSTMs, training can be slower than parallel models like transformers.
- **Difficulty with Very Long Sequences:** Although LSTMs can manage longer sequences than traditional RNNs, they still struggle with very long dependencies.

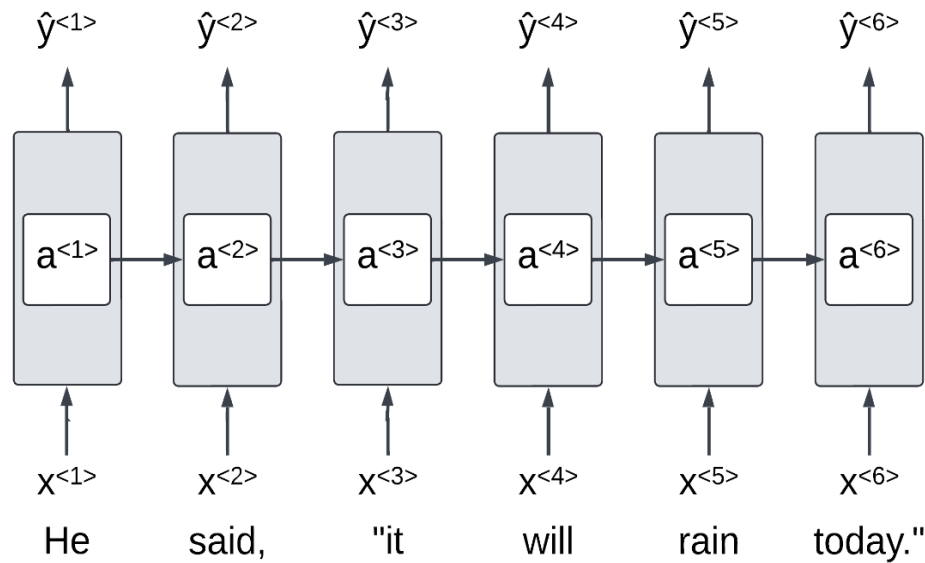


Figure 2 Architecture of an LSTM model

3.2.3. Bi-LSTM (Bidirectional LSTM)

Bi-LSTM is an extension of the standard LSTM that processes sequences in both forward and backward directions, capturing both past and future context. This bidirectional processing is crucial in tasks where the context before and after a word influences its meaning, such as fake news detection [5] [6] [16] [41].

Bi-LSTM consists of two LSTM layers: one that processes the sequence from start to end (forward direction) and one that processes it from end to start (backward direction). The outputs of both LSTMs are concatenated to form a richer representation of each word in the sequence [6] [16].

- **Input:** Sequential data (forward and backward directions).
- **Architecture:** Two LSTM layers (one for each direction), whose outputs are combined.

- **Output:** A contextual representation of words considering both previous and subsequent information.

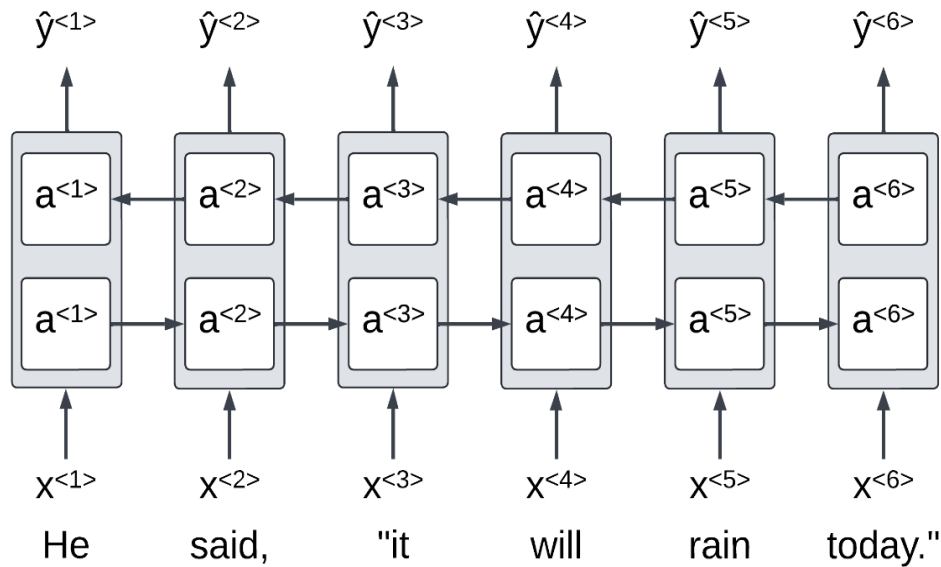


Figure 3 Architecture of a Bi-LSTM model

Pros:

- **Comprehensive Context:** By processing the sequence in both directions, Bi-LSTM captures more comprehensive contextual information.
- **Improved Performance:** It generally provides better accuracy in tasks where understanding both past and future context is essential.

Cons:

- **Computationally Heavier than LSTM:** Since it processes the data twice, in forward and backward directions, Bi-LSTM is slower and requires more memory compared to a unidirectional LSTM.

3.2.4. Random Forest

Random Forest is an ensemble learning method that operates by constructing multiple decision trees during training and outputting the mode of the classes (for classification) or mean prediction (for regression) of the individual trees. It helps in reducing overfitting and increasing the accuracy of the model [7] [31] [39].

Random Forest creates a multitude of decision trees and combines their results to make a more accurate prediction. Each tree is trained on a random subset of the data, and the final output is determined by aggregating the results from all trees (either by majority vote for classification or averaging for regression) [31] [39].

- **Input:** Random subsets of the dataset with selected features.
- **Architecture:** Multiple decision trees combined into an ensemble.
- **Output:** A class label or prediction based on the majority vote or average.

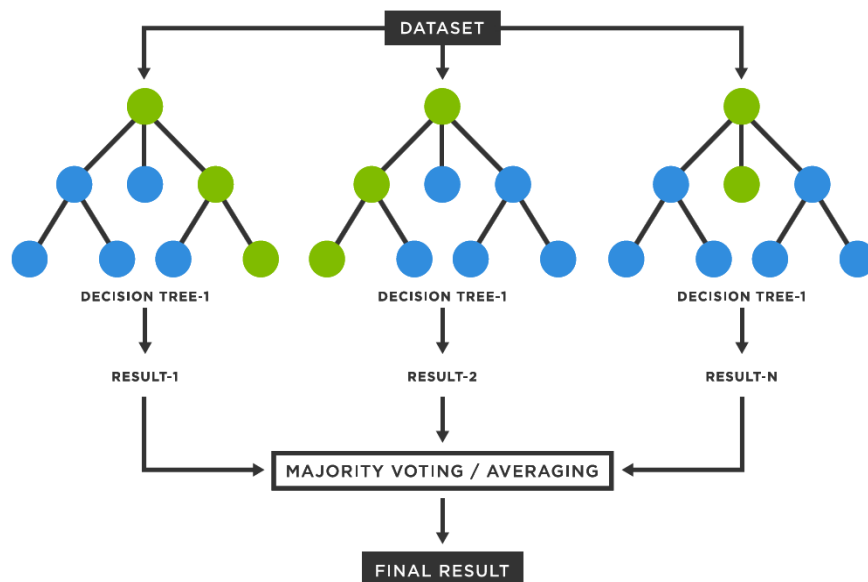


Figure 4 Architecture of a random forest model

Pros:

- **Reduces Overfitting:** By averaging the results from multiple trees, Random Forest avoids overfitting that is common with single decision trees.
- **High Accuracy:** The ensemble approach usually results in higher accuracy than single classifiers.
- **Handles Missing Data:** Random Forest is robust to missing data.

Cons:

- **Computationally Expensive:** Training and predicting with multiple decision trees can be resource-intensive.
- **Less Interpretability:** As the number of trees increases, it becomes difficult to interpret how the model arrives at its final prediction.

3.2.5. Decision Trees

A Decision Tree is a non-parametric model that splits the dataset into subsets based on the most important features, forming a tree structure. Each node represents a decision point based on a feature, and the branches represent the outcomes, leading to a final classification or regression at the leaves [28] [31] [39].

The decision tree uses metrics like **Gini impurity** or **entropy** to determine the best feature to split the data at each node. The tree grows recursively until all data points are correctly classified or until a stopping criterion is met (e.g., maximum depth or minimum number of samples at a node) [28] [31].

- **Input:** Feature vectors representing the dataset.
- **Architecture:** A tree-like structure where internal nodes represent feature splits, and leaf nodes represent class labels or regression values.
- **Output:** A predicted class or continuous value based on the decision rules learned during training.

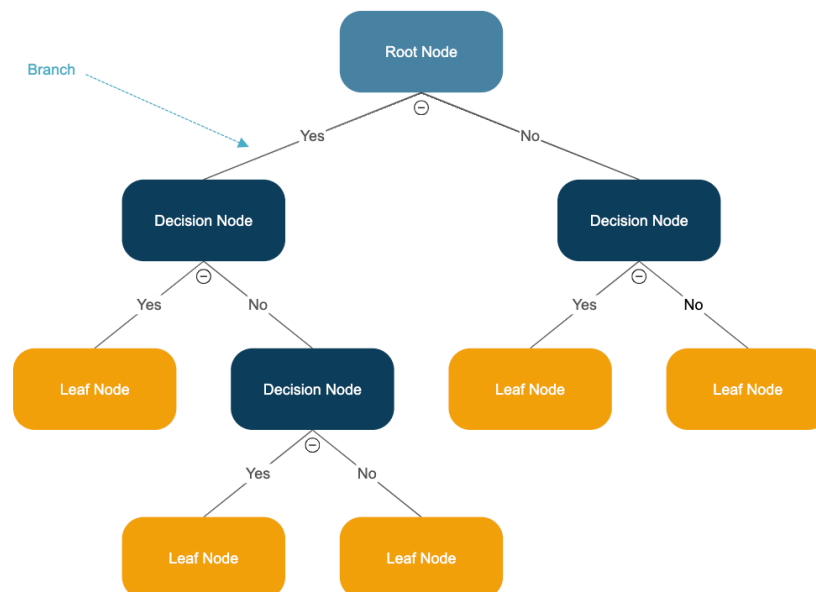


Figure 5 Architecture of a decision tree

Pros:

- **Simple and Interpretable:** The tree structure is easy to understand and visualize, making it interpretable.
- **Handles Categorical and Numerical Data:** Decision trees can handle both types of data well.

Cons:

- **Prone to Overfitting:** Without pruning or setting limits on tree depth, decision trees can easily overfit, especially on noisy data.
- **Bias with Imbalanced Data:** Decision trees can be biased toward dominant classes if the data is imbalanced.

3.2.6. Artificial Neural Network (ANN)

Artificial Neural Networks (ANNs) are inspired by the biological neural networks of the brain. They consist of layers of interconnected nodes (neurons) that are trained to recognize patterns in data. ANNs are highly versatile and can be used for a wide range of tasks, including classification and regression [1] [7] [10].

ANNs typically consist of an **input layer**, one or more **hidden layers**, and an **output layer**. Each neuron applies a non-linear activation function (such as ReLU or sigmoid) to the weighted sum of its inputs and passes the result to the next layer. The network is trained using **backpropagation**, where the error is propagated backward through the network to adjust the weights [7] [10].

- **Input:** Feature vectors representing the data.
- **Architecture:** Fully connected layers where each neuron is connected to every neuron in the next layer.
- **Output:** A class label or continuous value based on the task (classification or regression).

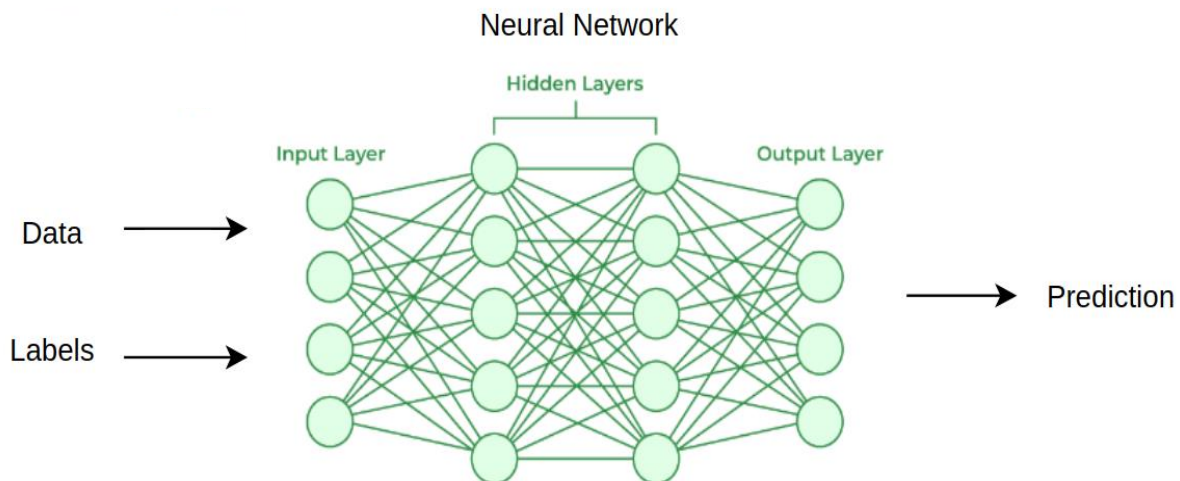


Figure 6 Architecture of the ANN model

Pros:

- **Capable of Modelling Complex Relationships:** ANNs can model non-linear and complex patterns in data, making them powerful for various tasks.
- **Scalability:** ANNs can be scaled to solve larger and more complex problems by increasing the number of layers and neurons.

Cons:

- **Requires Large Datasets:** ANNs generally require large amounts of data to perform well.
- **Black Box Nature:** It is difficult to interpret the learned representations in ANN models.

3.2.7. Support Vector Machine (SVM)

SVM is a supervised learning algorithm used for classification tasks. It works by finding the optimal hyperplane that maximally separates the data points of different classes. For non-linearly separable data, it can use the **kernel trick** to project the data into a higher-dimensional space where a linear separator can be found [9] [19] [20].

SVM tries to find the hyperplane that has the largest margin between the classes. The margin is defined as the distance between the hyperplane and the nearest data points of any class (called **support vectors**). SVM can handle both linear and non-linear classification tasks using different kernel functions (linear, polynomial, radial basis function (RBF)) [9].

- **Input:** Feature vectors representing the dataset.
- **Architecture:** A hyperplane in feature space, separating different classes.
- **Output:** A class label based on the position of the data point relative to the hyperplane.

Pros:

- **Effective for High-Dimensional Spaces:** SVM works well when the number of features is greater than the number of samples.
- **Robust to Overfitting:** Especially effective in cases where there is a clear margin of separation between classes.

Cons:

- **Computational Complexity:** Training an SVM, particularly with non-linear kernels, can be computationally expensive.
- **Less Effective on Large Datasets:** SVMs tend to scale poorly as the dataset size increases.

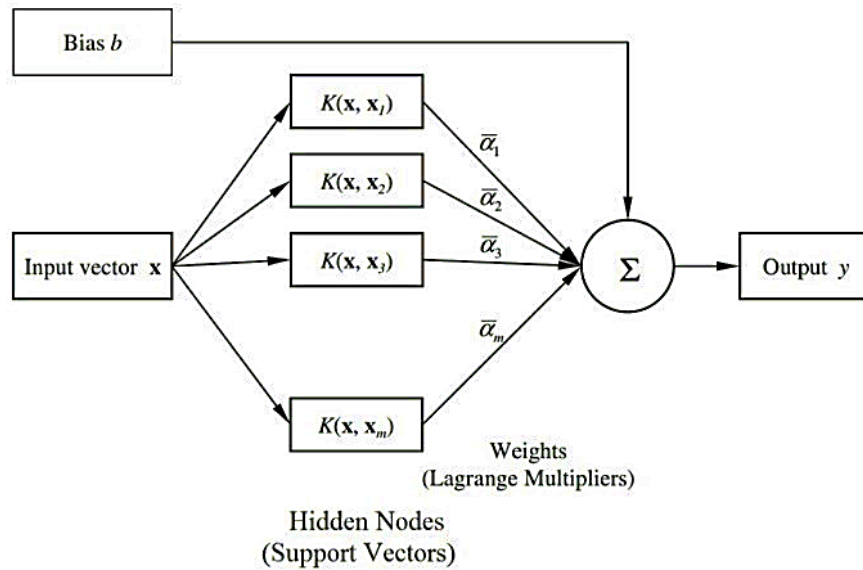


Figure 7 Schematic Diagram of SVM architecture

3.2.7. Logistic Regression

Logistic Regression is a linear model used for binary classification. It estimates the probability that a given input belongs to a particular class using the **logistic function**, which maps the output of a linear equation to a probability between 0 and 1 [7] [22] [28].

Logistic regression calculates the probability of an event by applying the logistic (sigmoid) function to a linear combination of input features. The model is trained by minimizing the **log loss** (cross-entropy), and the output is interpreted as a probability, which can then be thresholded to make a binary decision [7] [22] [34].

- **Input:** Feature vectors representing the dataset.
- **Architecture:** A linear equation followed by a logistic function.
- **Output:** A probability score for each class, typically used for binary classification tasks.

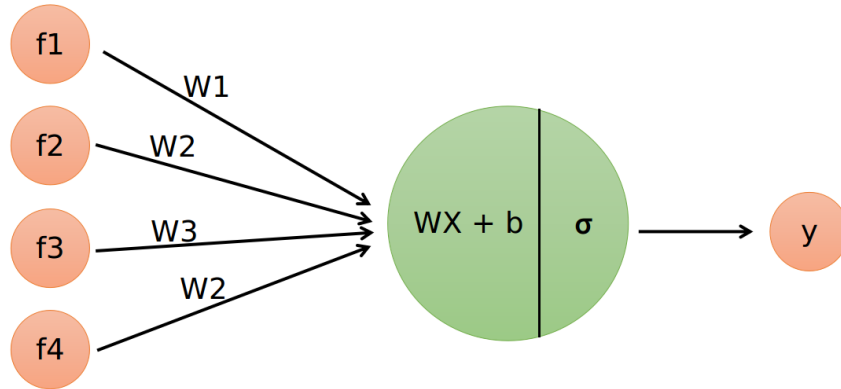


Figure 8 Architecture of the logistic regression model

Pros:

- **Simple and Interpretable:** Logistic regression is easy to implement and understand, making it a popular choice for initial modeling.
- **Fast and Efficient:** It works well for linearly separable data and is computationally efficient.

Cons:

- **Limited to Linear Boundaries:** Logistic regression assumes a linear relationship between the input features and the outcome, which can limit its effectiveness for complex, non-linear data.
- **Prone to Overfitting:** It can overfit when the dataset is small or when irrelevant features are present.

3.3. Proposed System Diagram

In this chapter, the datasets utilized for training, along with their statistical characteristics, are examined in detail. Subsequently, the various architectures employed for training and fine-tuning these datasets are discussed. Figure 9 presents the proposed system diagram, providing a

comprehensive overview of the processes involved. The following chapter will outline the integration of these components into the complete system.

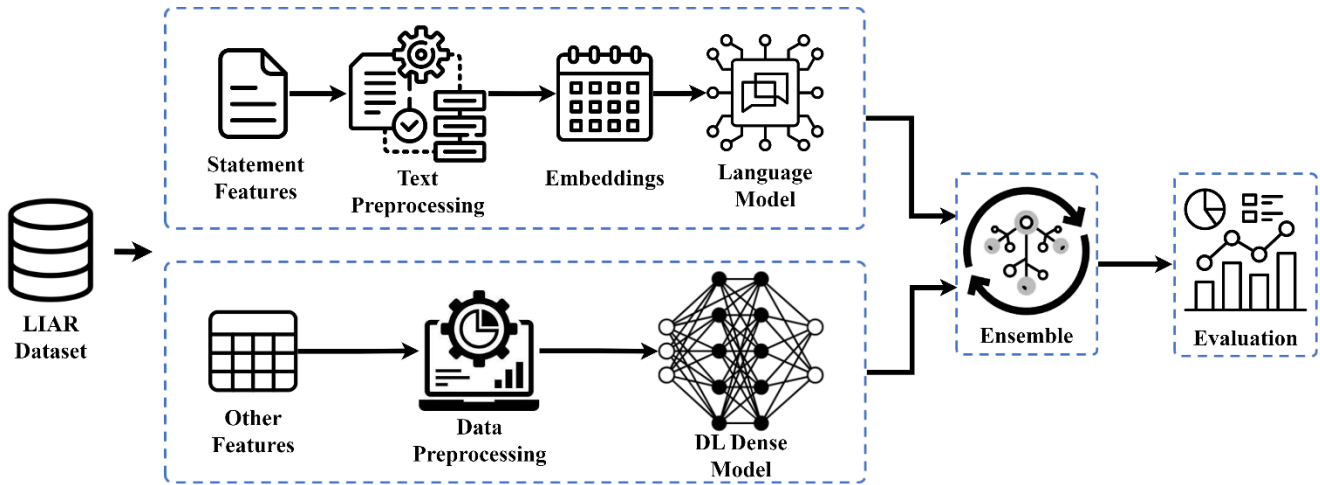


Figure 9 System architecture of the proposed hybrid model

CHAPTER 4: IMPLEMENTATION

This chapter outlines the process of developing an end-to-end automated system for detecting fake news. It includes a comparison of different models to identify the most suitable one for the task. Additionally, the chapter details the implementation of the solutions, covering the steps taken and the methodology used to achieve the objectives.

4.1. Data Collection and Preprocessing

In this section, we discuss data collection and the preprocessing techniques to prepare the data for model implementation.

4.1.1. Overview of the LIAR Dataset

The LIAR dataset is a widely recognized resource for fake news detection research [7]. It comprises 12,836 short news statements that have been manually labelled by fact-checkers at PolitiFact.com. These statements are attributed to 3,318 public speakers and come from various sources, including political debates, news releases, and social media posts. The average length of each statement is 17.9 tokens, making them concise yet information rich. The dataset’s comprehensive labelling and diverse topic coverage make it ideal for studying the complexity of real-world misinformation [7] [16] [33].

Each statement in the dataset is categorized into one of six truthfulness labels: True, Mostly True, Half True, Barely True, False and Pants on Fire [7].

Table 1: LIAR data statistics

Training set size	10,269
Validation set size	1,284
Test set size	1,283
Maximum words in a statement	467
Minimum words in a statement	4
Average tokens	17.9
# News	332
# Tweets	2589
# Verified users	550
# Unverified users	3,767
# Engagements	19,769
# Likes	5,713
# Retweets	10,434

Additionally, the dataset includes extensive metadata such as the subject of the statement, the speaker’s name, job title, political affiliation, context, and the sources of the information. This metadata is crucial for understanding the context in which the statements were made, adding depth to the analysis and improving the accuracy of fake news detection models.

Table 2: Data analysis

Int64Index	12791 entries, 0 to 1266		
Data columns	(total 14 columns):		
#	Column	Non-Null Count	Dtype
0	ID	12791 non-null	object
1	Label	12791 non-null	object
2	Statement	12791 non-null	object
3	Subject	12789 non-null	object
4	Speaker	12789 non-null	object
5	JobTitle	12789 non-null	object
6	StateInfo	12789 non-null	object
7	Party	12789 non-null	object
8	BarelyTrue	12789 non-null	object
9	False	12789 non-null	object
10	HalfTrue	12789 non-null	object
11	MostlyTrue	12789 non-null	object
12	PantsOnFire	12789 non-null	object
13	Venue	12789 non-null	object
memory usage	1.5+ MB		

4.1.2. Dataset Composition and Labelling

The **LIAR dataset** is meticulously organized, with each statement accompanied by a detailed verdict report from PolitiFact editors. This ensures that the labelling is not only accurate but also provides a comprehensive assessment of each statement’s truthfulness. The dataset was chosen for our research due to its ability to capture the nuances of misinformation across various contexts and topics.

Table 3: An example of the LIAR dataset

Statement	“During the recession, the consumer in his native perversity has begun to save. The savings rate is now 6.2 percent.”
Speaker Name	George Will
Current Job Position	Columnist
Home State	Maryland
Political Party	Columnist
Location of Speech	A roundtable discussion on ABC’s ”This Week”
Credit History	(7, 6, 3, 5, 1, 5)
Topic Label	Economy
Truthfulness Label	True

This example illustrates the depth of information provided for each statement, which includes not just the statement and its truthfulness but also contextual data such as the speaker’s credibility and the topic of discussion.

4.1.3. Dataset Statistics

The dataset is split into three subsets to facilitate model training, validation, and testing:

- **Training Set:** 10,269 entries
- **Validation Set:** 1,284 entries
- **Test Set:** 1,283 entries

These subsets ensure that models can be trained effectively, with separate data reserved for tuning and final evaluation. The distribution of words in statements varies significantly, with the shortest statement being **4 words** long and the longest reaching **467 words**. The average number of tokens per statement is **17.9**, highlighting the concise nature of the content [7].

4.1.4. Topic Distribution and Truthfulness

The LIAR dataset covers **144 news topics**, with the top 24 topics representing 80% of the total dataset. These topics include critical areas such as **economy**, **elections**, **healthcare**, and **immigration**, reflecting the dataset's focus on issues that are often subject to public debate and misinformation [7] [18] [33].

Table 4: Top 24 news topics

Economy	Health-care
Candidates-biography	Education
Elections	Federal-budget
Crime	Taxes
Immigration	Foreign-policy
Energy	Abortion
State-budget	Jobs
Guns	Campaign-finance
Children	Deficit
Congress	Corrections-and-updates
Environment	History
Job-accomplishments	Corporations

Figure 10 shows the distribution of truthfulness labels in the dataset:

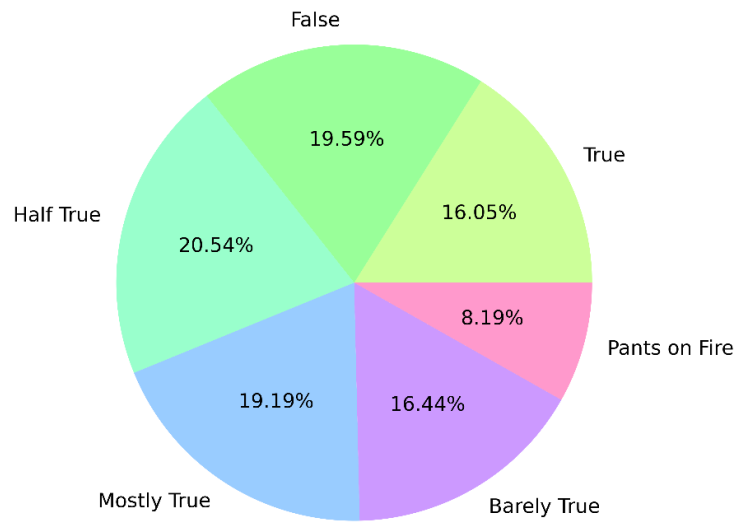


Figure 10 Distribution of Labels in LIAR Dataset

Figure 11 illustrates the distribution of statement lengths, with most statements being relatively short. The histogram shows that as the length of the statements increases, their frequency decreases sharply, indicating a skew towards brief statements. This introduces challenges in modeling, as shorter statements may lack context, making them harder to classify accurately.

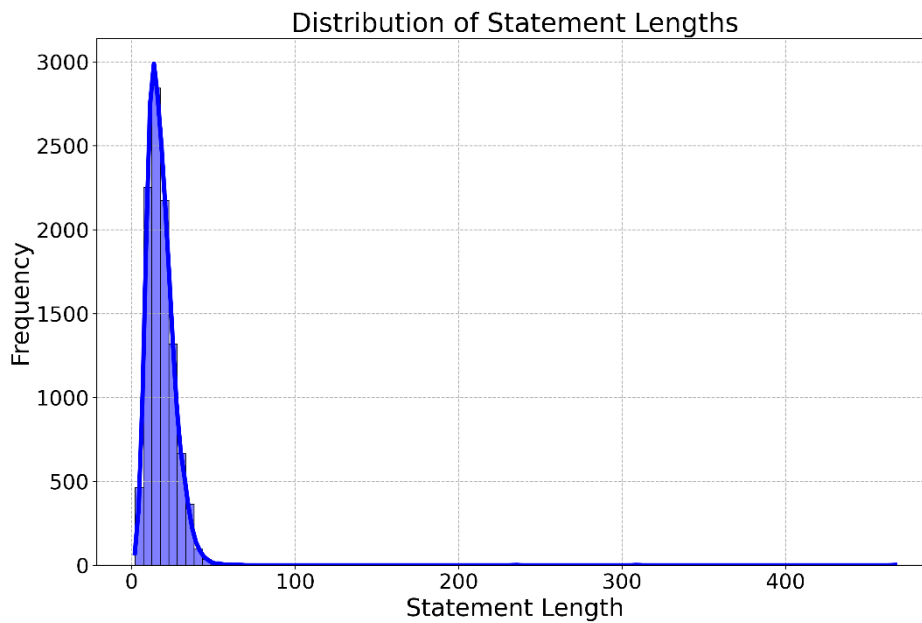


Figure 11 Distribution of statement lengths in the complete dataset

Figure 12 shows the distribution of statement lengths in terms of labels, highlighting the variance in statement complexity across different truthfulness categories.

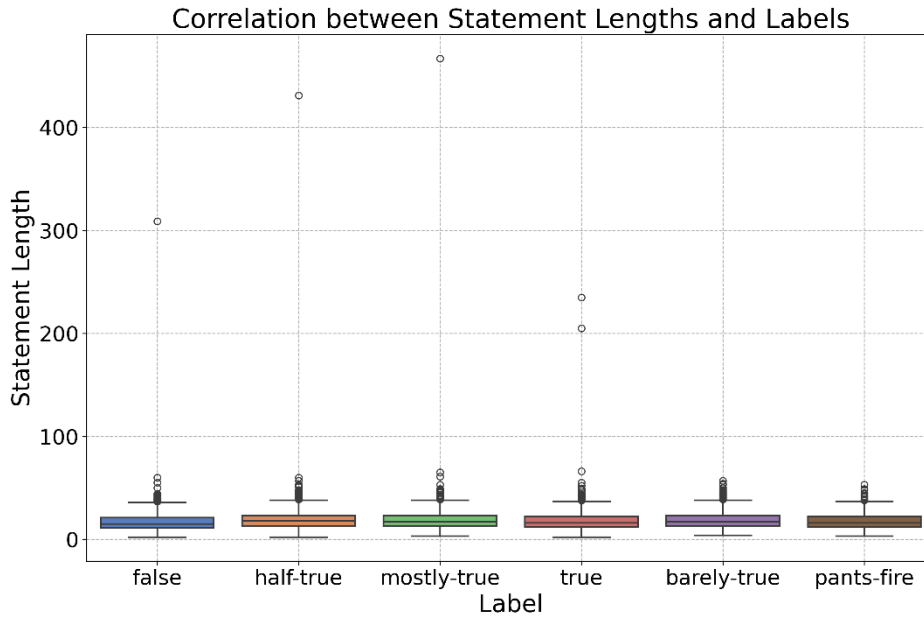


Figure 12 Distribution of statement lengths in terms of labels

Figure 13 displays the most frequent words in the dataset, providing insight into common themes and topics discussed in the statements.

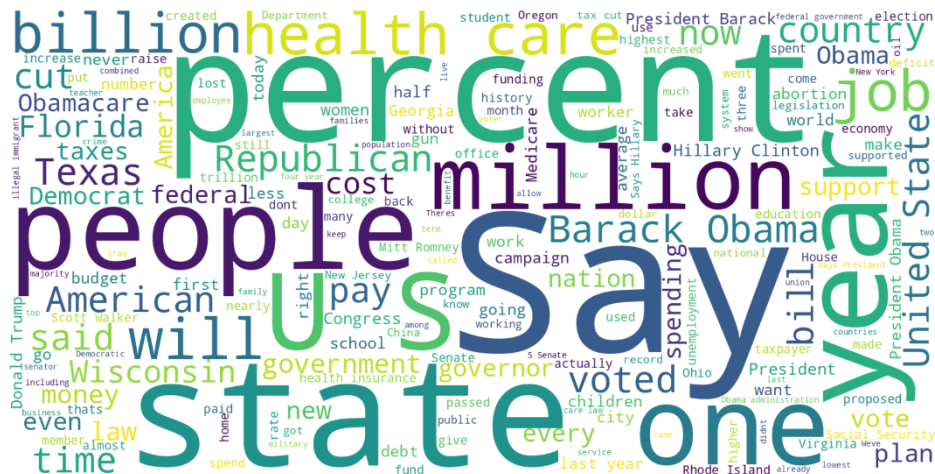


Figure 13 Most frequent words in the dataset

4.1.5. Challenges in Data Analysis

The LIAR dataset presents several challenges for analysis:

- **Imbalanced Categories:** The spread of truthfulness categories is uneven, with some labels being more prevalent than others. This imbalance can lead to biased model performance, where certain categories are predicted more accurately than others [7] [14] [33].
- **Short Statement Lengths:** The brevity of many statements makes it difficult for models to capture context and nuance, which are often crucial for accurate classification [7] [31].
- **Diverse Topics:** The wide range of topics covered in the dataset, from economy to healthcare, requires models to be versatile and capable of understanding various domains [7] [13] [19].

4.1.6. Label Definitions

The truthfulness labels used in the LIAR dataset are defined as follows:

1. **True:** The statement is correct, and nothing important is missing [7].
2. **Mostly True:** The statement is correct but requires additional information or clarification.
3. **Half True:** The statement is partially correct but omits key details or presents information out of context [7].
4. **Barely True:** The statement contains some truth but omits important details that would change the overall picture [7].
5. **False:** The statement is incorrect [7].
6. **Pants on Fire:** The statement is not only incorrect but also makes an outrageous claim [7].

These labels reflect the degree of accuracy of each statement, providing a nuanced view of truthfulness that goes beyond a simple binary classification of true or false [7].

4.2. Data Preprocessing Techniques

Effective data preprocessing is crucial for enhancing the performance of machine learning and deep learning models. In this research, a series of preprocessing steps were meticulously applied to the LIAR dataset to prepare the data for model training. Each step plays a vital role in ensuring that the data is clean, structured, and well-suited for analysis. Below, we discuss each

preprocessing technique in detail, highlighting its purpose, implementation, and impact on model performance.

4.2.1. Tokenization and Stopword Removal

Overview:

Tokenization is the process of converting a text into individual words or tokens, which are the basic units of data that a model can process. Stopword removal involves filtering out common words that do not contribute significant meaning to the text, such as "and," "the," "is," etc [25] [31] [33].

Implementation:

We utilized the **Natural Language Toolkit (NLTK)**, a powerful library for natural language processing (NLP), to tokenize the text from the LIAR dataset. Each statement was broken down into its constituent words. Following tokenization, stopwords were identified and removed from the dataset [16] [21].

Impact on Model Performance:

- **Efficiency:** By removing stopwords, we reduced the dataset's size and the computational resources required for model training. This streamlined the data, allowing the model to focus on the more meaningful words that contribute to the classification task.
- **Improved Accuracy:** Filtering out irrelevant words helps in reducing noise in the data, leading to better feature extraction and, ultimately, improved model accuracy.

4.2.2. Stemming

Overview:

Stemming is the process of reducing words to their root or base form. This technique helps in consolidating different forms of a word, ensuring that variations like "running," "ran," and "runner" are all treated as "run" [14] [23].

Implementation:

Stemming was applied to the tokenized words using NLTK's stemming tools. The stemming algorithm identifies the root form of each word and replaces the original word with its stem [21].

Impact on Model Performance:

- **Reduced Vocabulary Size:** Stemming significantly reduces the number of unique words (vocabulary size) in the dataset, which simplifies the model and reduces overfitting.
- **Better Generalization:** By grouping word variants under a single root form, stemming enhances the model's ability to generalize across similar words, improving its predictive accuracy.

4.2.3. Part-of-Speech (POS) Tagging

Overview:

POS tagging involves assigning labels to each word in a sentence based on its grammatical role, such as noun, verb, adjective, etc. Understanding the syntactical structure of a sentence through POS tags can provide valuable insights into the relationships between words [21] [28] [32].

Implementation:

Each word in the dataset was tagged with its corresponding POS label using NLTK. The top nine most frequent tags—*NOUN*, *VERB*, *ADP* (*adposition*), *PROPN* (*proper noun*), *PUNCT* (*punctuation*), *DET* (*determiner*), *ADJ* (*adjective*), *NUM* (*numeral*), and *ADV* (*adverb*) were assigned individual labels. Less frequent tags were grouped under a general label "X". Each POS tag was represented as a 10-dimensional one-hot vector, and a **10x10 identity matrix** was created, where each row corresponded to a specific POS tag embedding [32].

Impact on Model Performance:

- **Enhanced Feature Representation:** POS tags provide additional syntactical information that helps the model understand the grammatical structure and relationships between words in a sentence.
- **Improved Contextual Understanding:** By incorporating syntactical features, the model can better grasp the context and nuances of statements, leading to more accurate classification.

4.2.4. Clipping and Padding

Description:

Clipping and padding are techniques used to standardize the length of input sequences. In this study, each statement was clipped to a maximum length of 45 words. Statements shorter than this length were padded with zeros to ensure that all input sequences had consistent dimensions [14] [17].

Impact on Model Performance:

- **Consistency in Input Dimensions:** By standardizing the length of input sequences, clipping and padding allow the model to process batches of data more efficiently, leading to faster training times.
- **Reduced Overfitting:** Clipping helps to remove excessively long statements that may introduce noise, while padding ensures that shorter statements do not lose important context, both of which contribute to a more generalizable model.
- **Improved Model Stability:** Consistent input dimensions help prevent issues related to varying sequence lengths, ensuring that the model maintains stable performance across different batches of data.

4.2.5. Word Embeddings

Description:

Word embeddings are dense vector representations of words, capturing their semantic relationships in a continuous vector space. In this research, **GloVe** (Global Vectors for Word Representation) was used to generate 50-dimensional embeddings for each word in the statements [21] [23] [29].

Impact on Model Performance:

- **Semantic Understanding:** GloVe embeddings capture the semantic similarity between words, allowing the model to understand context more effectively, which is crucial for tasks like fake news detection.
- **Dimensionality Reduction:** By representing words as dense vectors, embeddings reduce the dimensionality of the input data, making the model more efficient without sacrificing the richness of the information.

- **Improved Generalization:** Pre-trained embeddings like GloVe are trained on large corpora, providing the model with a broad understanding of language that can improve performance even on unseen data.

4.2.6. Feature Extraction

Description:

Feature extraction involves converting text into numerical features that the model can process. In this study, **TF-IDF** (Term Frequency-Inverse Document Frequency) and **word2vec** techniques were used for feature extraction. TF-IDF captures the importance of words in a document relative to the entire dataset, while word2vec generates dense vector representations for each word based on its context [6] [15] [22].

Impact on Model Performance:

- **Enhanced Feature Discrimination:** TF-IDF helps the model focus on words that are more relevant to distinguishing between different classes, leading to better classification performance.
- **Contextual Embedding:** word2vec captures the context in which words appear, providing the model with richer, context-aware features that improve its ability to understand the nuances in statements.
- **Balanced Feature Representation:** The combination of TF-IDF and word2vec ensures that the model benefits from both frequency-based and context-based feature representations, enhancing overall accuracy.

4.2.7. N-grams

Description:

N-grams are contiguous sequences of n items (usually words) in a text. In this study, bi-grams (two-word sequences) and tri-grams (three-word sequences) were used to capture the context in which words appear, allowing the model to understand the relationship between words in a sequence [8] [11].

Impact on Model Performance:

- **Contextual Relationships:** N-grams help the model capture the relationships between words that appear close to each other, which is important for understanding idiomatic expressions, collocations, and phrases that may indicate fake news.
- **Enhanced Feature Set:** Incorporating n-grams into the feature set enriches the data, allowing the model to better understand the structure and meaning of statements, which is crucial for tasks like fake news detection.

4.3. Methodology

The methodology section of this research outlines the detailed process and approach employed to develop, train, and evaluate various models for fake news detection using the LIAR dataset. Our approach encompasses the use of several baseline models, a comprehensive architecture for the proposed hybrid model, and multiple machine learning and deep learning models, each tailored to capture different aspects of the data. The following sections provide an in-depth explanation of each component of our methodology.

4.3.1. Baseline Models

To ensure a comprehensive comparative analysis, we implemented three key baseline models that have significantly influenced subsequent research in the field of fake news detection. These models served as a foundation for evaluating the performance improvements achieved by our proposed hybrid model.

- **Basic LSTM Model:**

The Basic LSTM (Long Short-Term Memory) model is a standard sequential model without bi-directional processing or integration of syntactical features. This model processes input data sequentially, allowing it to capture temporal dependencies within the text. However, it does not consider context from both directions (past and future), which limits its effectiveness. Despite these limitations, the Basic LSTM model achieved an accuracy of 25.83% on the LIAR dataset, serving as a benchmark for more advanced models [4].

- **CNN Model:**

The Convolutional Neural Network (CNN) model was based on the architecture proposed by Wang (2017). CNNs are typically used in image processing but have been adapted for text classification tasks by applying convolutional filters to extract n-grams or local features from the text. This model achieved an accuracy of 31.76% on the LIAR dataset, demonstrating the effectiveness of CNNs in capturing local dependencies within text, although it falls short in capturing long-term dependencies [7].

- **Hybrid Model:**

The Hybrid Model combines Recurrent Neural Networks (RNNs) with user behavior analysis, as proposed by Ruchansky et al. (2019). This model integrates textual data with metadata such as user behavior, enhancing its ability to detect fake news by considering additional contextual information. The Hybrid Model achieved an accuracy of 36.94%, making it the most effective baseline model in our study. Its performance underscores the value of combining textual and non-textual data for more accurate classification [11].

Each baseline model was trained and evaluated using the same data preprocessing and training pipeline as our proposed model to ensure a fair comparison. The training process was conducted on Google Colab using a 15GB T4 GPU and 12.7 GB of RAM, providing sufficient computational resources for efficient model training.

4.3.2. Proposed Model Architecture

Our proposed model architecture is designed to leverage the strengths of both deep learning and ensemble learning techniques. It integrates a variety of features and processing techniques to enhance the model's ability to accurately classify fake news.

1. The Hybrid Model:

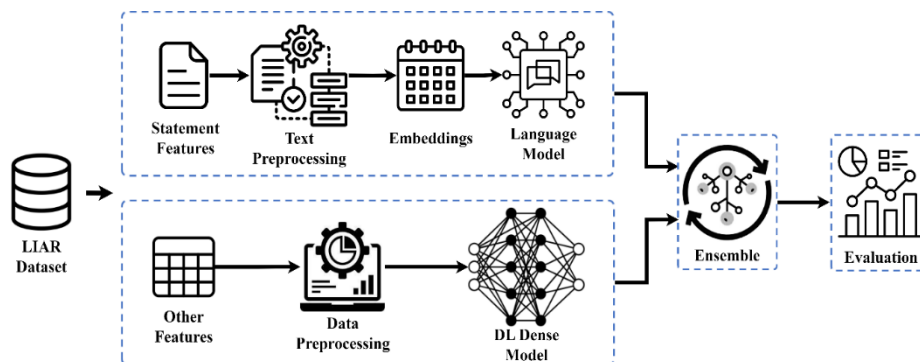


Figure 14 System architecture of the proposed hybrid model

- **Statement Features Processing:**
 - **Preprocessing:** Text from the LIAR dataset was tokenized, stopwords were removed, and stemming was applied to reduce words to their root forms. This preprocessing step prepares the text for efficient feature extraction by reducing noise and standardizing the input data.
 - **Embeddings:** Words were transformed into dense vectors using pre-trained embeddings like GloVe, which captures semantic relationships between words. These embeddings serve as the foundational layer for the model, providing rich, context-aware representations of the text.
 - **Language Model:** The embedded text was processed through a BERT (Bidirectional Encoder Representations from Transformers) model. BERT captures both contextual and syntactical details by considering the entire sentence structure and relationships between words. This step allows the model to understand the nuanced meaning of statements, crucial for accurately classifying fake news.

- **Other Features Processing:**
 - **Data Preprocessing:** Non-textual features, such as metadata (e.g., speaker information, context), were normalized and preprocessed to ensure compatibility with the model. These features are essential for providing additional context that can improve classification accuracy.
 - **DL Dense Model:** The preprocessed non-textual features were fed into a deep learning model consisting of three dense layers. These layers were designed to learn complex patterns within the data, allowing the model to capture intricate relationships that might not be apparent from the text alone.

- **Ensemble Learning:**
 - **Bagging (Bootstrap Aggregating):** Outputs from the language model (BERT) and the dense model (for non-textual features) were combined through ensemble learning. Bagging enhances predictive accuracy by aggregating the results of multiple models, leveraging the diverse feature sets to produce a more robust classification.

- **Evaluation:**

- **Performance Assessment:** The ensemble model was evaluated using standard metrics such as accuracy, precision, recall, and F1-score. These metrics provided a comprehensive view of the model's effectiveness across various categories, ensuring that the model performs well not only in overall accuracy but also in handling imbalanced data and nuanced truthfulness labels.

2. Random Forest:

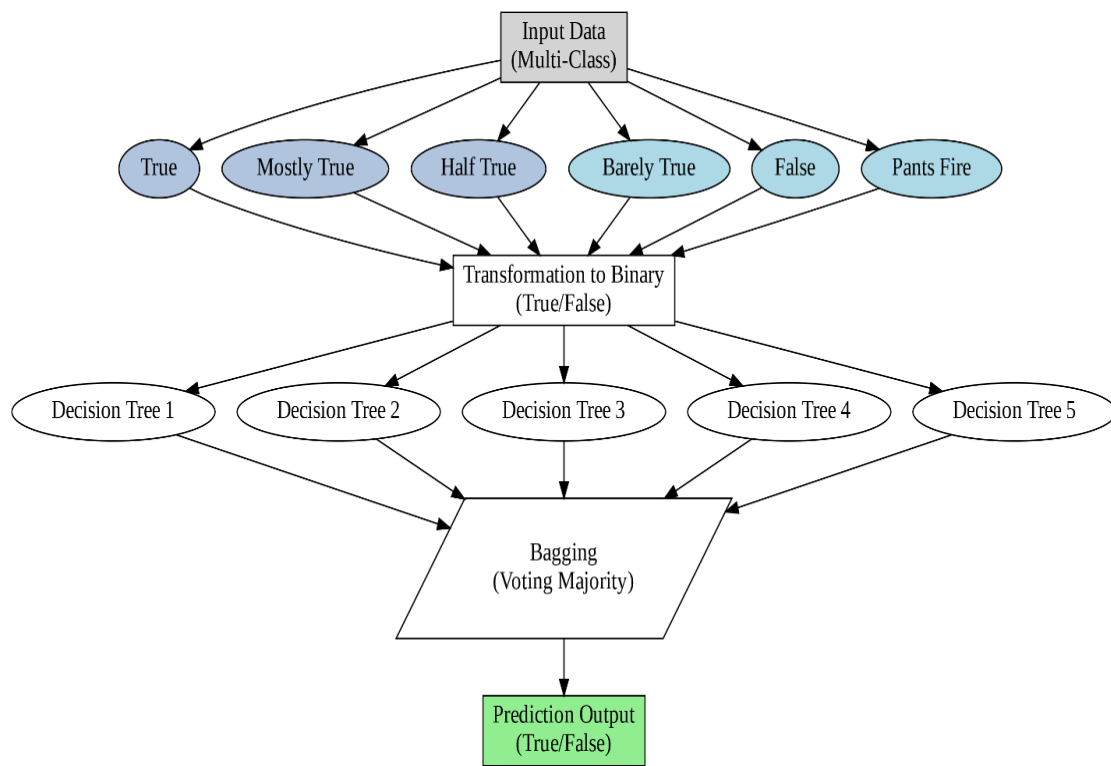


Figure 15 Flow diagram of the random forest for binary classification

- **Binary Classification Transformation:**

To simplify the multi-class problem, we transformed it into a binary classification task by merging the 'True', 'Mostly True', and 'Half True' classes into a single 'True' class, and the remaining three classes into a 'False' class. This transformation was aimed at improving model performance on specific tasks where binary classification is more effective.

- **Ensemble Learning with Random Forest:**

We implemented a Random Forest model, an ensemble learning method that combines multiple decision trees to improve classification accuracy. Random Forests are particularly effective at reducing overfitting, making them well-suited for this task. The model was also evaluated for six-way classification to provide a broader understanding of its capabilities.

3. **Bi-Directional Long Short-Term Memory (bi-LSTM) Network:**

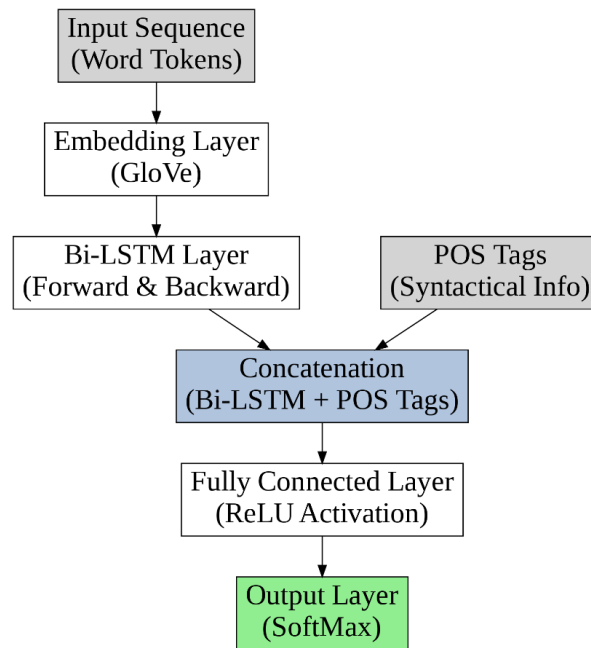


Figure 16 Flow diagram of LSTM for six-way classification

- **Embedding Layer:**

We used pre-trained GloVe embeddings to convert words into dense vectors, capturing the semantic relationships between words. These embeddings serve as the input to the bi-LSTM network.

- **Bi-LSTM Layer:**

Two bi-LSTM layers were incorporated into the network, each with 128 units. These layers process the input sequence in both forward and backward directions, capturing contextual

information from both preceding and succeeding words. Dropout and recurrent dropout of 0.2 were applied to prevent overfitting, enhancing the model's generalization capabilities.

- **POS Tag Integration:**

The output from the bi-LSTM layers was concatenated with Part-of-Speech (POS) tags, providing additional syntactical information that the model can use to improve its understanding of the text.

- **Fully Connected Layer:**

The concatenated features were passed through a dense layer with 32 neurons, using ReLU activation functions. This layer learns higher-level abstractions from the combined features, contributing to the final classification decision.

- **Output Layer:**

A SoftMax layer was used to output a probability distribution over the six truthfulness categories. The model was trained using the categorical cross-entropy loss function, optimizing it to correctly classify each statement into one of the six categories.

4. **Decision Trees:**

- **Implementation:**

A Decision Tree Classifier was implemented using entropy as the criterion for splitting. This model constructs a tree by recursively partitioning the data based on feature importance, allowing it to make decisions at each node.

- **Training and Evaluation:**

The Decision Tree model was trained on the dataset with a fixed random state for reproducibility. Decision Trees are highly interpretable and effective for classification tasks, particularly in identifying the most important features. However, they are prone to overfitting, which was managed using techniques like pruning and ensemble methods.

5. **Artificial Neural Network (ANN):**

- **Architecture:**

We implemented an ANN to capture non-linear relationships in the data, utilizing its capability to model complex patterns. The ANN consists of four dense layers with ReLU activation functions, providing flexibility and robustness in feature learning.

- **Training:**

The ANN was trained using backpropagation, a technique that iteratively adjusts the model's weights based on the error between predicted and actual outputs. This process ensures that the model learns from its mistakes, gradually improving its classification performance.

6. **Support Vector Machine (SVM):**

- **Implementation:**

SVM was implemented as it is effective in high-dimensional spaces and suitable for cases where the number of dimensions exceeds the number of samples. SVM models are particularly robust to overfitting, especially in high-dimensional space, making them ideal for our task.

- **Application:**

SVM was used both for binary and multi-class classification, providing a strong baseline for comparison with more complex models. Its effectiveness in handling linearly separable data added value to the overall methodology.

7. **Logistic Regression:**

- **Implementation:**

Logistic Regression was implemented for binary classification, leveraging its simplicity and effectiveness for linearly separable data. Like the Random Forest model, Logistic Regression was used both for binary and six-way classification, ensuring consistency across different classification tasks.

- **Training and Evaluation:**

The model was trained using a similar pipeline as the other models, ensuring that the results could be fairly compared. Logistic Regression serves as a simple yet effective benchmark, highlighting the improvements achieved by more advanced models.

4.3.3. Training and Optimization

The dataset was split into training, validation, and test sets using an 80-10-10 ratio. This split ensures that the models have sufficient data for training while also being evaluated on unseen data to measure generalization performance.

1. Data Preparation:

- **Sequence Representation:**

Each training data statement was reduced to 45 words, and each word was represented as a 50-dimensional vector from GloVe. This resulted in a statement embedding with a shape of (None, 45, 50). Additionally, 45 POS tags (one for each word in the statement) were generated, represented as 10-dimensional one-hot vectors, resulting in a POS-embedding with a shape of (None, 45, 10).

2. Optimization Techniques:

- **Early Stopping:**

Early stopping was employed based on the validation loss to prevent overfitting. This technique monitors the model's performance on the validation set and stops training when the performance no longer improves, ensuring that the model does not become too tailored to the training data.

- **Learning Rate Scheduler:**

A learning rate scheduler was used to dynamically adjust the learning rate during training, enhancing convergence. This technique helps the model to converge faster by adjusting the learning rate based on the progress of training.

- **Optimizer:**

The Adam optimizer was used to optimize the model. Adam combines the advantages of both AdaGrad and RMSProp optimizers, making it well-suited for tasks with sparse gradients and noisy data. The training process involved iterative forward and backward passes, updating the model weights to minimize the loss function.

3. Performance Metrics:

The models were trained and evaluated using standard performance metrics, including accuracy, precision, recall, and F1-score. These metrics provide a comprehensive view of the model's performance, ensuring that it is effective not only in overall accuracy but also in handling imbalanced data and nuanced truthfulness labels.

CHAPTER 5: RESULTS AND DISCUSSION

In this section, we thoroughly analyze the performance of our proposed hybrid model alongside other models implemented for both binary and six-way classification tasks using the LIAR dataset. The results presented offer a comprehensive comparison of various machine learning and deep learning models, demonstrating the effectiveness of our approach in enhancing the detection of fake news. Below is a detailed breakdown of the results and their implications.

5.1. Binary Classification Results

Binary classification is a critical task in fake news detection, where statements are classified as either true or false. The performance of the models on this task is summarized in Table V, which shows the accuracy and F1-scores for each model.

Table 5: Models' performances

Model	Accuracy	F1-score
Binary Classification		
LSTM	0.5417	0.5325
Random Forest	0.6249	0.6252
Logistic Regression	0.6018	0.5937
SVM	0.4394	0.4287
Six-way Classification		
Bi-LSTM	0.4173	0.4083
LSTM	0.3091	0.3115
Random Forest	0.6115	0.5913
SVM	0.3794	0.3710
Logistic Regression	0.6213	0.6244
ANN	0.2361	0.2303
Decision Trees	0.5230	0.5121
The Hybrid Model	0.6464	0.6420

5.1.1. LSTM:

- **Performance:** The Long Short-Term Memory (LSTM) model performed moderately well, achieving an accuracy of 54.17% and an F1-score of 53.25%.
- **Discussion:** LSTM models are well-suited for capturing temporal dependencies in sequential data, which is essential in understanding the flow of information in text.

However, the variability and complexity of fake news statements may have posed challenges, leading to less-than-optimal performance.

5.1.2 Random Forest:

- **Performance:** The Random Forest model outperformed other traditional models with an accuracy of 62.49% and an F1-score of 62.52%.
- **Discussion:** Random Forest's ensemble approach, which combines multiple decision trees, improves robustness and reduces overfitting, making it highly effective for binary classification tasks. This model's ability to handle various data features contributes to its superior performance.

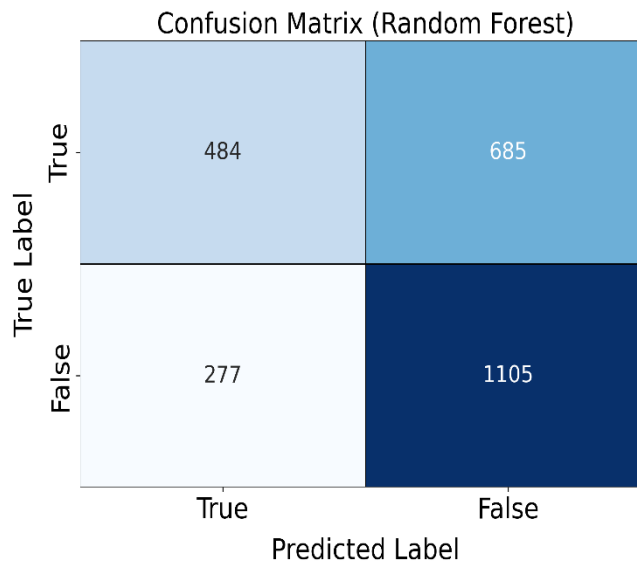


Figure 17 Confusion matrix for random forest in binary classification

5.1.3. Logistic Regression:

- **Performance:** Logistic Regression showed decent performance with an accuracy of 60.18% and an F1-score of 59.37%.
- **Discussion:** The simplicity of Logistic Regression, with its linear decision boundary, is effective in scenarios where the data is linearly separable. However, its performance may be limited in more complex datasets like the LIAR dataset, where non-linear relationships are present.

5.1.4. SVM:

- **Performance:** Support Vector Machine (SVM) lagged behind other models with an accuracy of 43.94% and an F1-score of 42.87%.
- **Discussion:** SVMs are typically powerful for high-dimensional spaces, but the linear kernel used may not have been sufficient to capture the underlying patterns in the data. The complexity of the fake news dataset, with its diverse and nuanced content, likely contributed to the model's lower performance.

5.2. Six-Way Classification Results

Six-way classification is a more challenging task, requiring the model to classify statements into one of six categories: True, Mostly True, Half True, Barely True, False, and Pants on Fire. Table V also presents the results for this task, highlighting the difficulties faced by different models.

5.2.1. Bi-LSTM and LSTM:

- **Performance:** Both Bi-LSTM and LSTM models exhibited lower performances, with accuracies of 41.73% and 30.91%, respectively.
- **Discussion:** The complex nature of multi-class classification, coupled with the varying lengths and complexity of statements, likely impacted the models' abilities to maintain high accuracy. Bi-LSTM, which captures contextual information from both directions, performed better than the standard LSTM, but both struggled with the nuanced nature of the data.

5.2.2. SVM and ANN:

- **Performance:** Both SVM and Artificial Neural Networks (ANN) performed poorly, with SVM achieving an accuracy of 37.94% and an F1-score of 37.10%, and ANN achieving an accuracy of 23.61% and an F1-score of 23.03%.
- **Discussion:** The SVM's difficulty in finding a suitable hyperplane for multi-class scenarios and ANN's potential insufficiency in depth or complexity likely contributed to their lower performance. These models struggled to capture the diverse patterns in the data, highlighting the need for more sophisticated approaches.

5.2.3. Random Forest and Decision Trees:

- Performance:** Random Forest and Decision Trees were among the better-performing traditional models, with Random Forest achieving an accuracy of 61.15% and an F1-score of 59.13%, while Decision Trees achieved an accuracy of 52.30% and an F1-score of 51.21%.
- Discussion:** Random Forest's ensemble nature allows it to manage the complexity of six-way classification by combining multiple decision trees, reducing variance, and enhancing predictive performance. Decision Trees, although simpler and more interpretable, struggled slightly with the complexity, reflected in their lower accuracy and F1-score. However, their ability to handle non-linear relationships and variable interactions remains valuable, especially when used in conjunction with ensemble methods.

Confusion Matrix (Random Forest)

	True	Mostly True	False	Half True	Pant's Fire	Barely True
True	95	108	67	80	10	35
Mostly True	88	203	100	84	6	34
False	110	133	115	122	9	45
Half True	75	96	96	146	4	52
Pant's Fire	53	70	36	33	8	17
Barely True	58	101	96	103	7	63
	True	Mostly True	False	Half True	Pant's Fire	Barely True
	Predicted Label					

Figure 18 Confusion matrix of random forest for six-way classification

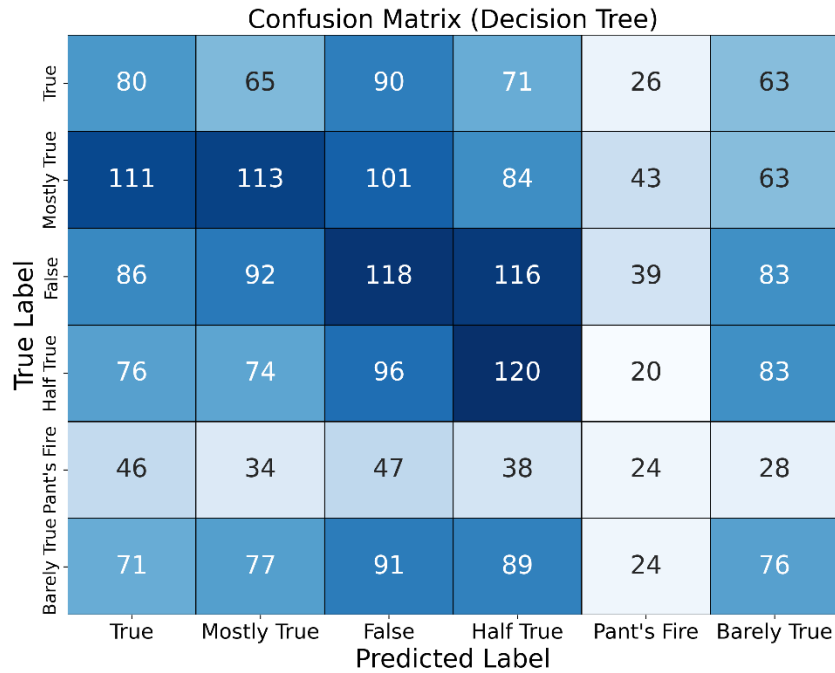


Figure 19 Confusion matrix of decision tree for six-way classification

5.2.4. Logistic Regression:

- **Performance:** Logistic Regression managed to achieve a relatively high accuracy of 62.13% and an F1-score of 62.44%.
- **Discussion:** Logistic Regression's simplicity and effectiveness in well-defined decision boundary scenarios likely contributed to its higher performance. Despite its linear nature, the model performed well in six-way classification, demonstrating its utility in certain contexts.

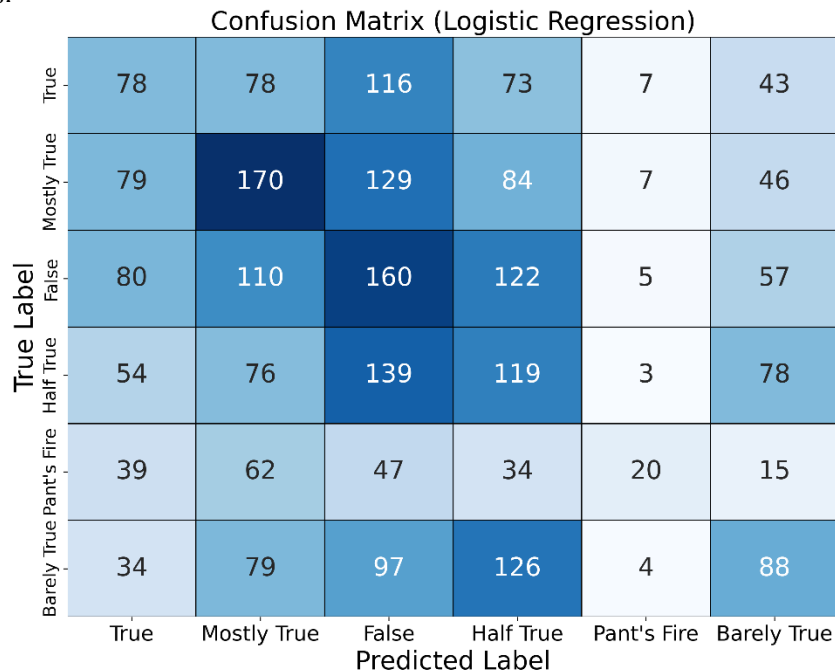


Figure 20 Confusion matrix of logistic regression for six-way classification

5.2.5. The Hybrid Model:

- **Performance:** Our Hybrid Model significantly outperformed all other models, achieving an accuracy of 64.64% and an F1-score of 64.20% in six-way classification.
- **Discussion:** The Hybrid Model's architecture, which integrates both textual and contextual features, leverages a combination of deep learning models and traditional methods to capture underlying patterns in both the content and context of statements. The ensemble of these features ensures that the model can generalize well, even in complex multi-class scenarios. The superior performance of our Hybrid Model emphasizes the importance of a multi-faceted approach to tackle the intricate problem of fake news detection.

5.2.6. Analysis of Confusion Matrix

The confusion matrix of our hybrid model, as shown in Figure 8, provides deeper insights into its performance across different categories. The following observations can be made:

Confusion Matrix (The Hybrid Model)

	Actual Labels	true	mostly true	half true	barely true	false	pants on fire	
	true	141	24	14	10	4	2	
	mostly true	20	167	19	18	12	9	
	half true	16	29	154	25	19	15	
	barely true	9	16	22	129	23	16	
	false	3	10	20	35	164	22	
	pants on fire	3	5	7	8	13	64	
		true	mostly true	half true	barely true	false	pants on fire	
		Predicted Labels						

Figure 21 Confusion matrix of the hybrid model for six-way classification

1. **Strong Performance in “Mostly True” and “Half True”:**

- The model demonstrated strong performance in identifying “mostly true” and “half true” statements, as evidenced by the high true positive rates in these categories.

2. **Misclassifications:**

- There were some misclassifications, particularly between “barely true” and “half true,” as well as “false” and “barely true” categories. This suggests that the model occasionally struggles to distinguish between closely related labels, likely due to the nuanced nature of these categories.

3. **Challenges with “Pants on Fire”:**

- The precision and recall for “pants on fire” were notably lower compared to other categories, indicating that the model finds it challenging to distinguish this extreme form of misinformation from other false statements. This difficulty could be attributed to the limited data samples available for this label, making it harder for the model to generalize.

Overall, the hybrid model's robust performance, particularly in comparison with other state-of-the-art techniques, highlights its effectiveness in fake news detection. The model's architecture, which effectively integrates both textual and contextual information, enhances its detection capabilities, particularly in handling complex and nuanced data.

5.3. Performance Comparison with Existing Models

To assess the effectiveness of our hybrid model, we compared its performance with several state-of-the-art methods applied to the LIAR dataset, as summarized in Table VI.

5.3.1. LSTM-Attention:

- **Performance:** This model uses a long short-term memory (LSTM) network with an attention mechanism that assigns varying importance to words, enhancing the model's focus on key terms crucial for fake news detection. However, the LSTM-Attention model exhibited a notable variance between precision and recall, particularly in the “pants on fire” label, where it achieved a precision of 0.514 but a recall of only 0.346 [13].

Table 6: Performance evaluation of fake news detection methods

Fake News Detection Model	Label	Precision	Recall	F1 Score	Accuracy	Macro-F1
LSTM-Attention [13]	True	0.373	0.165	0.229	0.390	0.402
	Mostly-True	0.432	0.424	0.428		
	Half-True	0.418	0.428	0.423		
	Barely-True	0.375	0.396	0.385		
	False	0.386	0.576	0.462		
	Pants-fire	0.514	0.346	0.413		
MMFD [32]	True	0.881	0.159	0.269	0.422	0.418
	Mostly-True	0.425	0.480	0.451		
	Half-True	0.322	0.636	0.427		
	Barely-True	0.586	0.290	0.388		
	False	0.453	0.479	0.466		
	Pants-fire	0.710	0.393	0.506		
Memory-Network [33]	True	0.916	0.156	0.267	0.452	0.449
	Mostly-True	0.421	0.483	0.450		
	Half-True	0.412	0.525	0.462		
	Barely-True	0.483	0.401	0.435		
	False	0.421	0.614	0.500		
	Pants-fire	0.661	0.509	0.575		
FDML [27]	True	0.567	0.564	0.565	0.508	0.516
	Mostly-True	0.493	0.530	0.511		
	Half-True	0.500	0.401	0.511		
	Barely-True	0.540	0.383	0.448		
	False	0.445	0.644	0.526		
	Pants-fire	0.662	0.554	0.604		
The Hybrid Model (Ours)	True	0.723	0.734	0.728	0.646	0.642
	Mostly-True	0.682	0.665	0.673		
	Half-True	0.600	0.653	0.625		
	Barely-True	0.601	0.573	0.590		
	False	0.650	0.698	0.673		
	Pants-fire	0.640	0.500	0.561		

5.3.2. MMFD:

- **Performance:** The Multi-Source Multi-Class Fake News Detection model introduces additional information from diverse sources, improving the model’s ability to differentiate between true and fake news across multiple categories. However, it also showed a significant variance between precision and recall, particularly in the “pants on fire” and “true” labels [32].

5.3.3. Memory-Network:

- **Performance:** Utilizing a memory network, this method emphasizes contextual information by integrating it as an attention factor. Despite its sophisticated approach, the

Memory-Network model struggled with consistency across labels, achieving high precision but low recall in some categories, such as the “true” label [33].

5.3.4. FDML:

- **Performance:** The Fake News Detection Multi-task Learning (FDML) model improves fake news detection by integrating representation learning with multi-task learning, focusing on both fake news detection and topic classification. While it showed improved results, it still struggled with categories like “false” and “barely true,” where there was a clear imbalance between precision and recall [27].

Our hybrid model demonstrated a significant improvement over these existing models, with approximately a 13.8% increase in average accuracy and Macro-F1 across the six truthfulness labels. The model's balanced scores across all labels, particularly excelling in the “true” label category, underscore its robustness and effectiveness in fake news detection.

CHAPTER 6: CONCLUSION AND FUTURE WORK

This chapter summarizes the results obtained from the current study. It also addresses the limitations of the proposed system and outlines potential future directions for further research.

6.1. Conclusion

In this research, we conducted a comprehensive examination of deep learning-based Natural Language Processing (NLP) techniques for detecting fake news using the LIAR dataset. Our focus was primarily on the development and evaluation of a hybrid model that integrates both textual and contextual features to enhance the accuracy and robustness of fake news classification. The hybrid model consistently outperformed existing models across a range of evaluation metrics, demonstrating its effectiveness, particularly in handling the nuanced and subtle truthfulness labels such as True, Mostly True, Half True, Barely True, False, and Pants on Fire.

The model's architecture, which leverages the power of Bidirectional Long Short-Term Memory (Bi-LSTM) networks alongside traditional machine learning methods, enabled it to capture the intricate relationships between words, context, and metadata. By incorporating part-of-speech (POS) tagging and dependency parsing, the model achieved a deeper syntactical and contextual understanding of statements. This additional layer of linguistic analysis significantly improved its performance, making it more capable of discerning subtle differences between similar statements, particularly when handling imbalanced data distributions.

Throughout the evaluation, the hybrid model demonstrated a balanced performance across all categories, consistently achieving higher accuracy, precision, recall, and F1 scores compared to traditional models like Logistic Regression, Decision Trees, Support Vector Machines (SVM), and other deep learning architectures such as basic LSTM and ANN. For instance, our hybrid model achieved an accuracy of 64.64% and an F1-score of 64.20% in six-way classification, marking a significant improvement over prior models applied to the LIAR dataset.

Moreover, this research has validated the importance of integrating both textual and contextual features for enhancing the detection capabilities of fake news models. The inclusion of metadata such as speaker credibility, political affiliation, and speech context proved to be highly valuable, enabling the model to offer a more comprehensive understanding of news authenticity. This is a vital step forward in mitigating the negative impacts of misinformation in today's fast-paced digital

environment. Additionally, our research offers a solid foundation for the development of real-time detection systems and multilingual models, addressing the growing diversity of online content.

6.2. Future Work

While this study has demonstrated significant advancements in the field of fake news detection, several areas of improvement and future research directions are worth exploring to further enhance model performance and applicability in real-world scenarios.

6.2.1. Enhancing the Model with Large Language Models (LLMs)

One of the most promising future directions is the integration of large language models (LLMs) such as GPT-4, T5, or BERT-based variants like RoBERTa or DeBERTa. LLMs have demonstrated superior performance in various NLP tasks due to their ability to capture context and semantics at a much deeper level [41] [44]. Integrating LLMs into the hybrid architecture could significantly improve the model's ability to understand long and complex statements, which are often present in news articles, political debates, and opinion pieces. The increased model depth and capacity could also enhance the detection of sophisticated fake news, such as those that manipulate factual information or use subtle linguistic cues to mislead the audience.

6.2.2. Incorporating Multilingual Capabilities

Fake news is a global phenomenon, spreading across various regions and languages. Currently, most fake news detection models, including ours, focus primarily on English datasets like LIAR. To extend the applicability of our model, future research could explore the incorporation of multilingual capabilities. This could involve training the model on multilingual datasets or leveraging translation models to convert non-English statements into English for classification. Additionally, applying LLMs capable of processing multiple languages natively could greatly enhance the model's ability to detect fake news across different linguistic contexts.

6.2.3. Real-Time Fake News Detection

Real-time detection is crucial for minimizing the spread of fake news, particularly on social media platforms where information dissemination happens almost instantaneously. Future research could focus on optimizing the model's computational efficiency and speed, making it capable of processing news articles, tweets, or posts in real-time. This might involve lightweight model

architectures or techniques such as model pruning, quantization, and distillation to reduce the computational load without compromising accuracy. Developing APIs or tools that can integrate with social media platforms for real-time analysis would further enhance the practical utility of this work.

6.2.4. Handling Data Imbalance with Advanced Techniques

Although our hybrid model performed well on the imbalanced LIAR dataset, further work is needed to explore more advanced techniques for handling data imbalance. Methods such as Synthetic Minority Over-sampling Technique (SMOTE), Generative Adversarial Networks (GANs) for data augmentation, or class rebalancing techniques could be employed to ensure that the model has sufficient exposure to minority classes like “Pants on Fire” and “True” labels [44] [45]. This would improve the model’s generalization ability, particularly for those categories with fewer samples.

6.2.5. Improving Explainability and Interpretability

As fake news detection models are increasingly applied in real-world scenarios, the need for explainability and interpretability becomes more important. Future work could focus on making the model’s decision-making process more transparent to ensure that users, particularly journalists and fact-checkers, can understand the rationale behind a model’s classification. Techniques such as SHAP (Shapley Additive Explanations) values or LIME (Local Interpretable Model-Agnostic Explanations) could be used to provide insights into which features or tokens were most influential in the model’s decision [43] [46].

REFERENCES

- [1] X. Zhou and R. Zafarani, “A survey of fake news: Fundamental theories, detection methods, and opportunities,” *ACM Computing Surveys (CSUR)*, vol. 53, no. 5, pp. 1–40, 2020.
- [2] A. Gelfert, “Fake news: A definition,” *Informal logic*, vol. 38, no. 1, pp. 84–117, 2018.
- [3] G. Pennycook and D. G. Rand, “The psychology of fake news,” *Trends in cognitive sciences*, vol. 25, no. 5, pp. 388–402, 2021.
- [4] R. Oshikawa, J. Qian, and W. Y. Wang, “A survey on natural language processing for fake news detection,” *arXiv preprint arXiv:1811.00770*, 2018.
- [5] H. Allcott and M. Gentzkow, “Social media and fake news in the 2016 election,” *Journal of economic perspectives*, vol. 31, no. 2, pp. 211–236, 2017.
- [6] W. Y. Wang, “‘liar, liar pants on fire’: A new benchmark dataset for fake news detection,” *arXiv preprint arXiv:1705.00648*, 2017.
- [7] X. Zhang and A. A. Ghorbani, “An overview of online fake news: Characterization, detection, and discussion,” *Information Processing & Management*, vol. 57, no. 2, p. 102025, 2020.
- [8] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake news detection on social media: A data mining perspective,” *ACM SIGKDD explorations newsletter*, vol. 19, no. 1, pp. 22–36, 2017.
- [9] D. Pogue, “How to stamp out fake news.” *Scientific American*, vol. 316, no. 2, pp. 24–24, 2017.
- [10] J. Y. Khan, M. T. I. Khondaker, S. Afroz, G. Uddin, and A. Iqbal, “A benchmark study of machine learning models for online fake news detection,” *Machine Learning with Applications*, vol. 4, p. 100032, 2021.
- [11] W. Shahid, B. Jamshidi, S. Hakak, H. Isah, W. Z. Khan, M. K. Khan, and K.-K. R. Choo, “Detecting and mitigating the dissemination of fake news: Challenges and future research opportunities,” *IEEE Transactions on Computational Social Systems*, 2022.
- [12] A. Kirilin and M. Strube, “Exploiting a speaker’s credibility to detect fake news,” in *Proceedings of Data Science, Journalism & Media workshop at KDD (DSJM’18)*, 2018.

- [13] Y. Long, Q. Lu, R. Xiang, M. Li, and C.-R. Huang, "Fake news detection through multi-perspective speaker profiles," in *Proceedings of the eighth international joint conference on natural language processing (volume 2: Short papers)*, 2017, pp. 252–256.
- [14] E. Essa, K. Omar, and A. Alqahtani, "Fake news detection based on a hybrid bert and lightgbm models," *Complex & Intelligent Systems*, vol. 9, no. 6, pp. 6581–6592, 2023.
- [15] J. George, S. M. Skariah, and T. A. Xavier, "Role of contextual features in fake news detection: a review," in *2020 international conference on innovative trends in information technology (ICITIT)*. IEEE, 2020, pp. 1–6.
- [16] F. T. Zuhra, K. Saleem, and S. Naz, "An accurate transformer-based model for transition-based dependency parsing of free word order languages," *Journal of King Saud University-Computer and Information Sciences*, p. 102107, 2024.
- [17] S. Hakak, M. Alazab, S. Khan, T. R. Gadekallu, P. K. R. Maddikunta, and W. Z. Khan, "An ensemble machine learning approach through effective feature extraction to classify fake news," *Future Generation Computer Systems*, vol. 117, pp. 47–58, 2021.
- [18] B. Bhutani, N. Rastogi, P. Sehgal, and A. Purwar, "Fake news detection using sentiment analysis," in *2019 twelfth international conference on contemporary computing (IC3)*. IEEE, 2019, pp. 1–5.
- [19] G. Curto, M. F. Jojoa Acosta, F. Comim, and B. Garcia-Zapirain, "Are ai systems biased against the poor? a machine learning analysis using word2vec and glove embeddings," *AI & society*, vol. 39, no. 2, pp. 617–632, 2024.
- [20] A. Tifrea, G. Bécigneul, and O.-E. Ganea, "Poincaré glove: Hyperbolic word embeddings," *arXiv preprint arXiv:1810.06546*, 2018.
- [21] M. J. G. Fagundes, N. T. Roman, and L. A. Digiampietri, "The use of syntactic information in fake news detection: A systematic review," *SBC Reviews on Computer Science*, vol. 4, no. 1, pp. 1–10, 2024.
- [22] A. Mitchell, J. Gottfried, M. Barthel, and E. Shearer, "The modern news consumer: News attitudes and practices in the digital era," 2016.
- [23] A. Willmore, "This analysis shows how viral fake election news stories outperformed real news on facebook," 2016.

- [24] N. Ruchansky, S. Seo, and Y. Liu, "Csi: A hybrid deep model for fake news detection," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, pp. 797–806.
- [25] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
- [26] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.
- [27] Q. Liao, H. Chai, H. Han, X. Zhang, X. Wang, W. Xia, and Y. Ding, "An integrated multi-task model for fake news detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 11, pp. 5154–5165, 2021.
- [28] I. Agarwal, D. Rana, K. Panwala, R. Shah, and V. Kathiriya, "Analysis of contextual features' granularity for fake news detection," *Multimedia Tools and Applications*, vol. 83, no. 17, pp. 51835–51851, 2024.
- [29] N. Aslam, I. Ullah Khan, F. S. Alotaibi, L. A. Aldaej, and A. K. Aldubaikil, "Fake detect: A deep learning ensemble model for fake news detection," *complexity*, vol. 2021, no. 1, p. 5557784, 2021.
- [30] K. Shu, S. Wang, and H. Liu, "Exploiting tri-relationship for fake news detection," *arXiv preprint arXiv:1712.07709*, vol. 8, 2017.
- [31] M. H. Goldani, R. Safabakhsh, and S. Momtazi, "Convolutional neural network with margin loss for fake news detection," *Information Processing & Management*, vol. 58, no. 1, p. 102418, 2021.
- [32] H. Karimi, P. Roy, S. Saba-Sadiya, and J. Tang, "Multi-source multi class fake news detection," in *Proceedings of the 27th international conference on computational linguistics*, 2018, pp. 1546–1557.
- [33] T. T. Pham, "A study on deep learning for fake news detection," 2018.
- [34] Arkaan, Shabiq Ghazi, Aldy Rialdy Atmadja, and Muhammad Deden Firdaus. "Fake news detection in the 2024 Indonesian general election using Bidirectional Long Short-Term

- Memory (BI-LSTM) algorithm." *Fake news detection in the 2024 Indonesian general election using Bidirectional Long Short-Term Memory (BI-LSTM) algorithm* 21.2 (2024): 22-30.
- [35] Zhang, Qin, et al. "A deep learning-based fast fake news detection model for cyber-physical social services." *Pattern Recognition Letters* 168 (2023): 31-38.
- [36] Farhangian, Faramarz, Rafael MO Cruz, and George DC Cavalcanti. "Fake news detection: Taxonomy and comparative study." *Information Fusion* 103 (2024): 102140.
- [37] Kuntur, Soveatin, et al. "Fake News Detection: It's All in the Data!." *arXiv preprint arXiv:2407.02122* (2024).
- [38] Madani, Mirmorsal, Homayun Motameni, and Reza Roshani. "Fake news detection using feature extraction, natural language processing, curriculum learning, and deep learning." *International Journal of Information Technology & Decision Making* 23.03 (2024): 1063-1098.
- [39] Alnabhan, Mohammad Q., and Paula Branco. "Fake News Detection Using Deep Learning: A Systematic Literature Review." *IEEE Access* (2024).
- [40] Alghamdi, Jawaher, Yuqing Lin, and Suhuai Luo. "Unveiling the hidden patterns: A novel semantic deep learning approach to fake news detection on social media." *Engineering Applications of Artificial Intelligence* 137 (2024): 109240.
- [41] Peng, Liwen, et al. "Not all fake news is semantically similar: Contextual semantic representation learning for multimodal fake news detection." *Information Processing & Management* 61.1 (2024): 103564.
- [42] Abualigah, Laith, et al. "Fake news detection using recurrent neural network based on bidirectional LSTM and GloVe." *Social Network Analysis and Mining* 14.1 (2024): 40.
- [43] Mallik, Abhishek, and Sanjay Kumar. "Word2Vec and LSTM based deep learning technique for context-free fake news detection." *Multimedia Tools and Applications* 83.1 (2024): 919-940.
- [44] Alghamdi, Jawaher, Suhuai Luo, and Yuqing Lin. "A comprehensive survey on machine learning approaches for fake news detection." *Multimedia Tools and Applications* 83.17 (2024): 51009-51067.
- [45] Zhao, Wenbin, et al. "Fake News Detection Based on Knowledge-Guided Semantic Analysis." *Electronics* 13.2 (2024): 259.

- [46] Zeng, Fanhao, et al. "A Multimodal Knowledge Representation Method for Fake News Detection." *2024 4th International Conference on Computer, Control and Robotics (ICCCR)*. IEEE, 2024.