



Fake News Detection Using Machine Learning

Prepared By

Name: [Metab Hussain Jdeba]

IDNo.: [252001230]

Submitted in fulfillment of the requirements for the degree
Master of Science in Artificial Intelligence

At

College of Informatics

Midocean University

UAE

Supervisor Name: Dr. Ashraf El-Sayed

College of Informatics, Midocean University, Comoros

[2025-semester2]

Acknowledgements

I would like to express my sincere gratitude to Dr. Ashraf El-Sayed for his invaluable guidance, continuous support, and encouragement throughout this research project. His expertise, insightful feedback, and unwavering dedication have significantly contributed to the quality and direction of this thesis.

I also extend my heartfelt appreciation to my family, friends, and colleagues, whose patience, motivation, and moral support were instrumental in the successful completion of this work.

Abstract

The accelerated development of the online media and social networking platforms has transformed the information sharing process and, at the same time, established the environment which can be prone to the distribution of misinformation and fake news. Such a phenomenon is very dangerous to the opinion of people, the democratic process, and stability in society, and one of the urgent challenges of the world today is to detect fake news. The manual verification and fact-checking can never identify the vast amount and pace of the information spreading in the digital age.

The study will present a machine learning model of automatically identifying fake news through textual analysis. The model is based on preprocessing and converting raw text into structured format using natural language processing to ensure it is used in computational learning. Linguistic patterns will be represented in numbers with the help of such feature extraction methods as Bag-of-Words and Term Frequency-Inverse Document Frequency (TF-IDF). A range of supervised machine learning algorithms such as Logistic Regression, Support Vector Machine (SVM), Random Forest, and Naive Bayes will subsequently be designed and tested to establish their capability of classifying the content of news as real or fake.

Standard evaluation measures, including the accuracy, precision, recall, F1-score, and ROC-AUC are going to be used to evaluate the performance of the proposed models. It is predicted that a strong, interpretable scaled machine learning system can be developed to detect the fake news articles with high accuracy. By offering information on the weaknesses and strengths of classical methods of machine learning, this study will be aimed at contributing to the literature on automated misinformation detection.

List of Publications

Table of Contents

Acknowledgements	ii
Abstract.....	iii
List of Publications	iv
List of Abbreviations	vii
Table of Figures	viii
Table of Tables	ix
Chapter 1 Introduction	1
1.1. Introduction	1
1.2. Problem Background	2
1.3. Problem Specification and Motivation.....	3
1.4. Research Questions & Objectives.....	5
1.4.1. Research Questions.....	5
1.4.2. Research Objectives.....	6
1.4.3. Expected Contribution	7
1.5. Research Significance	8
1.6. Thesis Structure.....	9
Chapter 2 Overview	11
2.1. Introduction	11
2.2. Conceptual Introduction and Fake News definition.....	12
2.3. Representation of Features and Data Modalities.....	13
2.4. Conventional machine learning practices.....	14
2.5. Deep Learning Approaches	15
2.6. Multi-View, Graph, and Multi-Modal Models.....	15
2.7. Domain and Language-Specific Studies.....	16
2.8. Surveys, Comparative Studies and Problems.	17
2.9. Summary and Research Gaps.....	18
2.10. Chapter Summary	22
Chapter 3: Research Methodology.....	24

3.1. Introduction	24
3.2. Dataset Description	24
3.3. Data Preprocessing.....	26
3.4. Model Development.....	27
3.5. Evaluation Metrics	29
3.6. Research Roadmap	30
3.7. Summary	32
References.....	33

List of Abbreviations

Abbreviation Full Term

AI	Artificial Intelligence
AE	Autoencoder
BiLSTM	Bidirectional Long Short-Term Memory
BoW	Bag-of-Words
CNN	Convolutional Neural Network
CSV	Comma-Separated Values
DL	Deep Learning
F1-Score	Harmonic Mean of Precision and Recall
GCN	Graph Convolutional Network
GBCA	Graph-Based Co-Attention
ISOT	University of Victoria Information Security and Object Technology (Fake News Dataset)
KNN	K-Nearest Neighbors
LM	Language Model
LSTM	Long Short-Term Memory
ML	Machine Learning
MV-AE	Multi-View Autoencoder
NLP	Natural Language Processing
PUI-Net	Post-User Interaction Network
RNN	Recurrent Neural Network
ROC-AUC	Receiver Operating Characteristic – Area Under the Curve
SVM	Support Vector Machine
TF-IDF	Term Frequency–Inverse Document Frequency
XAI	Explainable Artificial Intelligence

Table of Figures

Figure 1 Experimental Workflow	7
Figure 2 visual representation of how the chapters, research files, and model iterations interact throughout the study	10
Figure 3 Preprocessing stages	26
Figure 4 Model Architecture	28
Figure 5 Work Flow	28

Table of Tables

Table 1 Summary of Key Studies in Fake News Detection Using Machine Learning	21
Table 2 breakdown of the dataset categories	25
Table 3 Research Roadmap.....	31

Chapter 1 Introduction

1.1. Introduction

The online news consumption has overtaken the traditional media in the digital age and this has resulted in the proliferation of information exponentially. Although this new digital world has made information more democratic to people, there is an increasing circulation of fake news and information that challenge public trust, stability in politics and social harmony. The fast spreading of fake news by means of social networks, news portals, and short video platforms has led to a critical necessity of the automated and intelligent detection systems that will be able to identify the fabricated or manipulated information in real-time.

Detection Fake news has become an important subdiscipline of Machine Learning (ML) studies. Different researchers have conducted researches to compare content-based and context-based methods of differentiating authentic and counterfeit news articles. Pereira et al. [1] continued this trend by introducing multi-view autoencoder networks, which can be used to train deep models to learn cross-domain representations to learn semantic subtleties between heterogeneous data views.

Some survey studies have summarized the developments and problems in the field. The methods of multi-view bootstrapping presented by Ying et al. [2] also improve the detection strength by using various news features, including author credibility, and propagation dynamics.

New attempts have been focused on the exploitation of user interaction networks and credibility modeling. Min et al. [3] suggested a divide and your conquer model that characterizes the user interaction on the social media to achieve higher accuracy in fake news detection whereas Bazmi et al. [4] devised a co-

attention network that encodes the topic specific user interaction and the credibility of the news source.

The growing sophistication of the misinformation patterns has prompted deployment of deep learning structures and representation learning models. Zhang et al. [5] integrated Graph Convolutional Networks (GCN) and the models of BERT-based co-attention, which are important in enhancing the semantic reasoning of detection models.

Multilingual misinformation has also become a challenge of interest. Alkudah et al. [6] conducted a comparative study of machine learning and deep learning frameworks on fake news detection in Arabic media by showing that hybrid solutions that integrate CNN and Bi-LSTM are better than the traditional classifiers, particularly when they are trained on language-specific data.

Although there have been major improvements, the area of fake news detection continues to have unresolved problems, which include dataset bias, model interpretability, domain adaptation, and cross-lingual scalability.

1.2. Problem Background

With the introduction of the internet and the overwhelming growth of social media sites, the spread of information is immediate and international. This revolution has shifted the very nature of news consumption among the people in preference to speed over fact-checking. Although this connectivity can be used to share knowledge quickly, it can also be used to propagate fake news which can be defined as intentionally made or misleading information about a specific topic being presented as factual content [1]. This kind of misinformation has been identified to distort the perceptions of the people, affect campaigns, fuel social unrests, and this has also led to a lack of trust in professional journalism [5], [6].

Early detection models were very dependent on superficial linguistic features namely word frequency, sentiment polarity and grammatical structure, to differentiate between genuine and fake news [1], [6]. Nevertheless, the development of fake news material has rendered these strategies inadequate. The modern information distortions involve the use of sophisticated language structures, emotionally charged stories, and multi-modal features (images, videos, hashtags), which have to be comprehended by more sophisticated models that can discern a semantic, contextual, and relational pattern [2], [1]. In this regard, scholars have investigated numerous paradigms. Content models examine semantics of the text, whereas context models combine user interaction information and source reliability [2], [4]. Combining textual, visual, and social dimensions, hybrid and multi-view learning strategies, like the one suggested by Pereira et al. [1] can enhance the effectiveness of detection.

Though these have been advanced, there are still a number of challenges. First, there is a tendency to have dataset imbalance and domain dependency thus resulting in models, which do well in one application environment, and fail in a different application environment [6]. Second, semantic ambiguity and intentional sarcasm or satire make the detection through text more difficult [5]. Third, the emergence of AI-generated content and deep fake videos has brought about novel types of misinformation, which cannot be reduced to text manipulation [2].

1.3. Problem Specification and Motivation

The main issue that will be considered in the study is automated detection and categorization of fake news articles by applying computational techniques which could be applied to a variety of topics and writing styles. Although there is a lot of news data online, it is not a trivial task to differentiate between valid and fake information such is the advanced level of misinformation production and

distribution. The authors of fake news consciously construct the language of the articles, the emotional feelings, and the structure of the article as authentic to resemble real journalism [1], [6]. Due to this, machine learning systems will require the ability to acquire nuanced semantic and contextual associations that can not be explained by superficial lexical patterns.

The statement of the problem can be therefore expressed as follows:

Having a set of labeled news articles with text, title, subject and date of publication, the objective will be to develop and test machine learning models that can distinguish between real and fake news articles in terms of the text and structural characteristics of the article. The system is expected to work with raw text using a clear pipeline that consists of the data cleaning, tokenization, feature extraction (e.g., TF-IDF or word embeddings), and supervised learning algorithms. To compare the performance of various models, standard evaluation measures will be used which include accuracy, precision, recall, F1-score, and ROC-AUC.

The aim behind this issue is the increasing social significance of misinformation and the necessity to have large-scale, automated solutions that can alleviate its effects. NLP methods are efficient to capture meaningful linguistic features to detect fake news, and in contrast, Ying et al. [2] show that the multi-view or hybrid learning models are significantly more efficient in terms of accuracy and adaptability to the domain.

Practically, fake news is a multidimensional problem in that it affects elections, poses a threat to the health of the population, and divides societies [5], [3]. As an example, in times of a crisis on a global scale, including the case of pandemics or conflicts, factual information can be disseminated slower than misinformation, which will decrease institutional trust and civic adherence. The initial point of defense could be automated detection systems, which will alert

about potentially harmful content before it gets into mass audiences. Also, the growing popularity of AI-created media, including large language models and text-generating applications, requires that counter-AI models be created that can identify artificially created fake information [10].

1.4. Research Questions & Objectives

The prevalence of misinformation on the digital platform has rendered fake news detection a highly important investigation field in machine learning (ML). To overcome this issue, the current research develops a system of research questions and objectives, which will be used to develop, implement, and assess an automated fake news detector system.

1.4.1. Research Questions

The key questions guiding this research are as follows:

- RQ1: What is the way to effectively integrate machine learning to identify fake news articles based on textual information?
- RQ2: Which preprocessing and feature extraction methods (e.g., Bag-of-Words, TF-IDF, Word2Vec) are the most effective in terms of their high accuracy and generalization in a wide range of news topics?
- RQ3: Which machine learning or deep learning algorithm, like Logistic Regression, Random Forest, Support Vector Machine (SVM), or Long Short-Term Memory (LSTM), is the most trustworthy in terms of classification in the detection of fake news?
- RQ4: How effective are state-of-the-art multi-view and explainable AI solutions in improving model interpretability and trust in fake news detection systems?
- RQ5: What are the ways that the results of this research can lead to the development of scalable, transparent and domain adaptive fake news detectors that can be deployed in the real world?

Such research questions provide both technical and ethical aspects of an issue-technical in how to create effective algorithms, and ethical in how to make automated misinformation filtering transparent and accountable [1], [6], [10].

1.4.2. Research Objectives

The growing complexity of the production and dissemination of fake news has posed pressing concerns about the manner in which computational systems can be able to differentiate authoritative and fake information correctly and ethically. In this regard, the proposed research will seek to understand how to incorporate Machine Learning (ML) tools to develop the powerful and explainable model that can identify fake news. The main research question is motivated by the fact that it is necessary to comprehend how text-based and semantic attributes can be identified and exploited to enhance the extent to which deceptive content can be classified. Therefore, some research questions have been developed to inform this study.

The research purposes are clearly associated with the questions provided above and are aimed at developing a multifaceted framework of fake news detection based on text data. Particularly, the research aims at preprocessing and cleaning the data set through the method of tokenization, removal of stop-words and lemmatization to warrant the quality and uniformity of textual input. Moreover, it will attempt to derive meaningful linguistic and semantic content in the news articles with the help of vectorization techniques like TF-IDF and word representations, make the text information to be effectively represented in terms of computational models. The study also aims to design and train several of them, such as Logistic Regression, Random Forest, Support Vector Machine, and Long Short-Term Memory (LSTM) to assess and compare the results of their usage in detecting fake news. The standard performance measures including accuracy, precision, recall, F1-score, and ROC-AUC will be used to evaluate the models as they present a broad picture of the ability of any particular model to classify data.

These objectives are driven by two things, academic and societal. Academically, the study has been able to enhance the current advancement of the computational methods used in detecting misinformation by providing the benchmark between the traditional and deep methods on a large balanced dataset. Socially, the research contributes to the fight against the proliferation of misinformation in the entire world and rebuilding digital trust in news media.

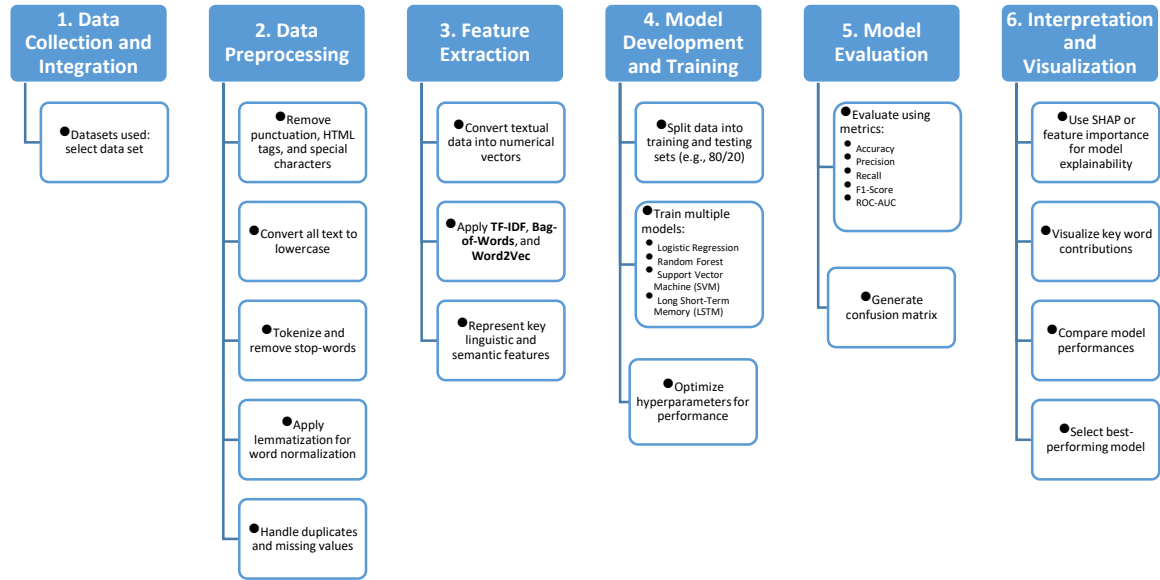


Figure 1 Experimental Workflow

1.4.3. Expected Contribution

By answering the above objectives, the study will also be able to make a contribution to the growing field of work in computational misinformation detection through:

- Setting up a benchmark system to classify fake news on a structured English data set.
- Comparison and contrast of classical and neural methods of textual misinformation analysis.
- Emphasizing the need to make machines explain why something is true or false and the need to be ethical in content verification systems.

Overall, the proposed study aims to address issues related to the gap between theoretical research and real-life application to suggest a model that is structured, data-driven, and explainable and can help media organizations, policymakers, and technology platforms detect and counteract the spread of fake news in real-time.

1.5. Research Significance

This study is important because it contributes to the current efforts of the world to fight misinformation by using the latest machine learning (ML) methods. The uncontrolled spread of fake news has become a serious social, political, and economic danger in the age where the consumption of digital content prevails. The misinformation may be used to control the minds of the people, disrupt the democracies, and make people make decisions on various matters including health and finances. Most of the manual fact-checking procedures, however useful, cannot cope with the sheer volume of data that is produced day in, day out on the online platforms. Thus, creating automated, data-driven and scalable systems that can detect counterfeit news is an academic requirement as well as a social need.

Scientifically, the current study is important to the academic discussion since it presents a systematic analysis of the detecting models of fake news based on the Fake News Detection Dataset of True.csv and Fake.csv. Through the combination of several text preprocessing, feature extraction, and classification techniques, this study does not only compare the performance of the classical ML models, including Logistic Regression and Random Forest, but also explores the possibilities of deep learning structures, like LSTM. This comparative methodology improves the knowledge of the influence of linguistic and semantic representations on the model performance in discriminating between factual and fabricated news.

At the practical level, the results of this research can be used to establish the basis of development of real-life applications that can monitor and filter online news streams on live basis. These systems can assist journalists, policymakers, and social media platforms to detect and reduce misinformation prior to their spreading to large masses. Besides, the implications of the findings on the design of AI-assisted fact-checking software and teaching systems that may foster media literacy and critical information consumption among users are presented. This study hopes to bring the divide between research and practical application by helping the whole world move towards information integrity, accountability, and digital trust.

1.6. Thesis Structure

The thesis is as follows:

- Chapter 1 presents history, statement of the problem, purpose and importance of the research.
- In chapter 2, the literature review on fake news detection from multiple perspectives, including linguistic-based, network-based, and multimodal approaches.
- Chapter 3 provides description of the proposed methods such as data pre processing, graph modeling, and machine learning and deep learning models employed.
- Chapter 4 writes about the experimental configuration, experimental results and analysis, presenting the performance outcomes of each model and comparing their effectiveness in detecting fake news.
- Chapter 5 is a conclusion of this work that throws emphasis on the important findings, limitations, and future directions of the study.

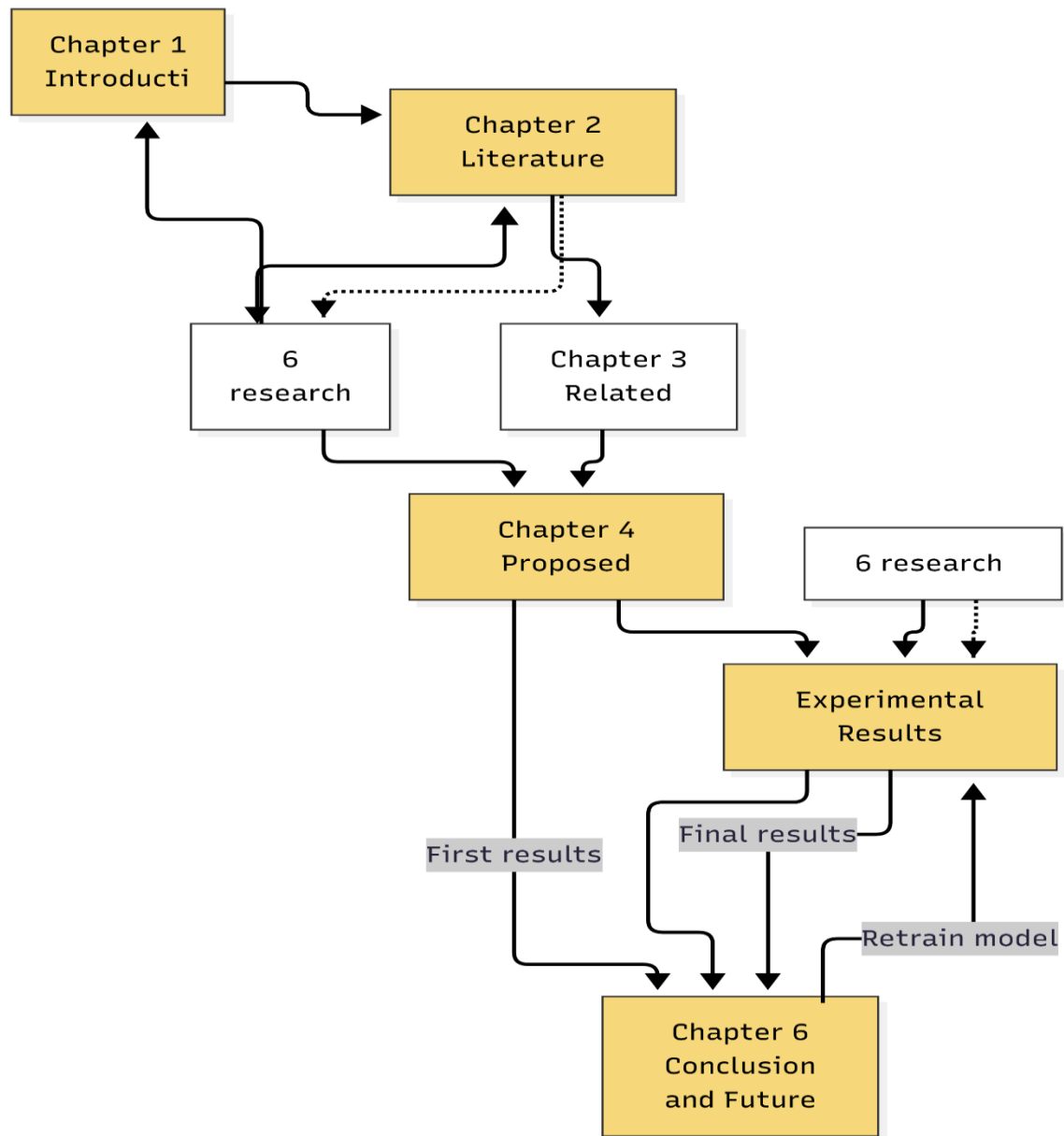


Figure 2 visual representation of how the chapters, research files, and model iterations interact throughout the study

Chapter 2 Overview

2.1. Introduction

Fake news has become one of the most topical phenomena of the digital information age that has significant impacts on the opinion of people, political rhetoric, and social harmony. Because of the popularity of social media sites and the ability to publish online, fake information can now reach out to millions of users in seconds, which can be verified by traditional means, notwithstanding. As a result, scientists and technologists have more and more concluded to the use of computational intelligence, or to be more precise, machine learning (ML) and deep learning (DL) to automate the process of misleading or deceptive news content. This chapter is the review of the literature on the topic of the fake news detection with the help of the ML and DL approaches, analyzing the conceptual background, methodological tendencies, and technological innovations that influenced the given sphere.

The research on fake news detection can be divided into some broad streams. The former stream consists of content-based techniques, which examine textual or linguistic features e.g. syntax, semantics and sentiment with the aim of revealing deceptive patterns. The second is context based and social based in approach with a strong interest in the role of user behavior, propagation networks, and source credibility in differentiating between authentic and fabricated information. The third and the latest stream incorporates multi-view and multi-modal frameworks which are a combination of different perspectives textual, visual and social in order to understand the complex and dynamic nature of misinformation dissemination.

Machine learning techniques, such as Support Vector machines (SVM), random forest and Logistic regression, have long dominated initial research because they are interpretable and also efficient in structured feature spaces.

2.2. Conceptual Introduction and Fake News definition.

The fake news has become a widely spread notion of the digital age as the misinformation spreads faster on the online media platforms. In its most basic definition, fake news can be called the deliberate creation or deception of information which is offered under the guise of real news in order to elicit certain ideological, political, or financial implications on the audience. It is not just a question of textual falsity but it is a concerted mechanism of dissemination on both sides, on the one hand, content producers and on the other hand, social network players [4].

Analytically speaking, initial studies concentrated mainly on the detection of linguistic anomalies, i.e., vocabulary use, mood, composition, etc. as a form of falsification. The later research, however, identified that the intent of deception and virality could not be completely captured with the help of text because users and credibility of the source are also important factors [9]. It is this development that gave rise to hybrid analytical paradigms that combine natural language processing, network analysis and behavioral psychology.

The fake news conceptual definition also changed with the technological progress. With neural networks and language models reaching the point where they could produce human-like text [5], the distinction between genuine and synthetic text became even less distinct, and more advanced detection mechanisms are required. Modern definitions, therefore, include intentional falsification as well as automated generation of misinformation and this time might not necessarily be human authored.

2.3.Representation of Features and Data Modalities.

The representation of features is a crucial aspect of fake news detection as the ability of a machine learning system to separate authentic and fake content depends on it. It is always stressed in the literature that a credible representation should reflect not only the language features of the news text but also the social and cultural aspects of its distribution. Content based representations dwell on the textual characteristics of the news articles that include lexical patterns, semantic coherence, and sentiment polarity [5].

Conversely, features based on the context are dependent on external features, including the level of user interaction, credibility of the source, and the pattern of propagation through social networks. The paper by Min et al. [3] coined the notion of post-user interaction networks, showing that the structural data regarding the user interaction, including comment threads or retweet graphs, can be much more predictive than textual information. In the same manner, Bazmi and Shakery [4] have modeled user and source credibility on topics by adopting multi-view co-attention models and highlighted that reliability of a claim is closely related to the trustworthiness of the claim publisher and the audience.

The representations that have become popular include hybrid and multi-view representations since they combine various modalities into a single latent space. Ying et al. [2] introduced a textual and social views bootstrapped multi-view strategy which concurrently undertakes textual and social views to enhance multi-platform generalization. Pereira et al. [1] developed this direction by introducing multi-view autoencoders which are trained on the heterogeneous data sources that learn shared representations, effectively integration of linguistic information, user behavior and visual information[5].

Together, these works reflect the paradigm shift of a single-modality analysis of the text to a multi-modal feature integration, where the text, user behavior, visual

or auditory content are in interaction with each other. This development is an indication of an increased realization of the fact that misinformation detection involves a comprehensive grasp of what is being expressed, how, and by whom it is expressed. These full-fledged representations are the foundations of the modern deep learning architectures and will guide the future generation of fake news detectors[6],[7].

2.4. Conventional machine learning practices.

Prior to the emergence of deep learning, fake news identification was entirely dependent on conventional machine learning (ML) algorithms, which implied the need to perform explicit feature engineering. These traditional methods had a basis in the belief that fake news could be detected by the help of the specific linguistic and statistical features. According to Kesarwani et al. [7], the algorithms initially used in the social media misinformation detection were the K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Naive Bayes, Decision Trees, and Logistic Regression. These models rely on hand designed features based on word frequencies, part-of-speech labels, sentiment polarity and stylistic differences.

The traditional ML methods have one benefit in that they are interpret-able; feature weights and decision boundaries give understandings into the linguistic and structural cues of deception. Along with these strengths, the conventional ML models have significant drawbacks [7]. They rely on handcrafted features, which are sensitive to variations in domains and languages besides being time consuming to extract. Such models find it hard to be generalized in data-sets due to the fact that fake news writing styles vary in regions and subjects. Moreover, they are not easy to learn even simple contextual or temporal interactions, like how misinformation propagates over user interaction networks, without significant feature prediction [8].

2.5. Deep Learning Approaches

The development of deep learning has had a significant impact on the evolution of fake news detection since it substituted the manual feature engineering with data-driven representation learning. Contrary to the classical structure of machine learning models, deep architectures are capable of automatically learning high-dimensional correlations between the linguistic features, semantics and contextual information [10].

Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) and their variants, like the Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) models are among the most popular architectures that are widely used. The RNN-based architectures were used to identify fake news by Cetiner [9], and it was established that sequential modeling of linguistic patterns can allow the systems to learn how to associate the time dependencies between sentences.

The GBCA model proposed by Zhang et al. [5] combines Graph Convolutional Networks (GCN) with BERT based on the co-attention line, and it has been shown to yield better generalization when combining language and relationship features. These architectures are effective to learn representations of the text content and the network relations between the users and publishers.

Multi-modal information integration has also been made easy through deep learning models. Even though deep networks have recorded state-of-the-art results, they exhibit a black-box behavior that presents challenges on interpretability and explainability [10].

2.6. Multi-View, Graph, and Multi-Modal Models.

The recent years have seen the growing interest in multi-view, graph-based and multi-modal models to detect fake news, as the idea that misinformation cannot be perceived through the text messages only has taken hold. Rather it is born out

of the interaction of various views - textual, social, temporal and visual - which are used to give different signals. Pereira et al. [1] proposed multi-view auto-encoders (MV-AE), where the latent representations are trained with the help of encoding a variety of information streams such as the text of the article, social context, and user interaction data. This unsupervised method is superior to single-view baselines and finds association between different sources that are heterogeneous. On the same note, Ying et al. [2] suggested a bootstrapped multi-view representation system to match textual and propagation embedding and hence overcome the issue of partial or inaccurate data metrics by enabling one view to complement the other.

It has turned into a building block of future detection systems that is based on graphs. A multi-view co-attention network proposed by Bazmi and Shakery [4] also focuses on topic-specific relationships between users and news sources, and credibility as a relationship feature, but not as an intrinsic attribute. This can be best achieved with Graph Convolutional Networks (GCNs) by modeling the diffusion of fake news over social ties and diffusion of trust or skepticism over groups of users.

The multi-modal architectures further extend the paradigm by incorporating text with other modalities like images, audio and video. Zhang et al. [5] introduced a co-attention mechanism (GBCA) to unite GCN and BERT that allows learning semantics of texts and network structure simultaneously. These models also indicate a significant epistemic change: fake news is ceased to be viewed as the document classification task but as a multi-relational phenomenon that is dynamic [6].

2.7.Domain and Language-Specific Studies.

Even though most of the research regarding the detection of fake news is conducted using English language datasets, recent research indicates the

significance of the model adaptation to different linguistic and cultural contexts. The idiomatic expressions, the language peculiarities, and the dynamics of social media play a major role in spreading and perceiving misinformation. One of the most thorough studies on the detection of Arabic fake news was done by Alkudah et al. [6], who compared the applications of machine learning and deep learning to the Arabic Fake News Dataset. Their results showed that although deep learning models (e.g., CNNs and LSTMs) are more effective than traditional models, linguistic preprocessing is important because of the morphological complexity of Arabic and dialect differences. Equally, Al-Taie [9] conducted a comparative study in which word embeddings specific to the Arabic language (e.g., AraVec and multilingual BERT) have a significant positive effect on the performance of detection, as opposed to generic embeddings.

The problem of domain adaptation is not only limited to language but also to content type. As an example, author [3] concentrated on short-video media where misinformation is presented in visual and auditory, as opposed to textual, form.

2.8. Surveys, Comparative Studies and Problems.

The development of the fake news detector has been a well-documented phenomenon with a significant number of surveys and comparative studies, all of which provide a systematic perception of the methods, data, and issues. One of the most detailed taxonomies is introduced author [10], who divide ways of detecting fake news into content-based, context-based, and hybrid frameworks. They compared the effectiveness of the linguistic and the network-driven ones and have found that the hybrid models, which incorporate the user interaction, publisher credibility, and semantic coherence, have always been more efficient than one view system only. Their discussion indicated the overlap of fact-

checking and fake news detection especially in the field of natural language inference and knowledge-graph reasoning.

Domenico et al. [8] conducted a systematic review and associated fake news, social media dynamics, and marketing, showing how fake news campaigns take advantage of the emotion contagion and the algorithmic amplification processes. Like these insights help to connect the technical and behavioral approaches to the problem and support its interdisciplinary nature.

Relative comparative studies like those conducted by Alkudah et al. [6] tested how machine learning and deep learning models were relative to each other in the Arabic setting. Their findings affirm that linguistic diversity and an unbalanced dataset are important in accuracy and present a great limitation of cross-domain generalization.

There are a number of common challenges in all comparative works: labeled multilingual data is scarce, the models can be adversarial manipulated, there are no standard benchmarks, and the models are challenging to explain. Also, the majority of the detection systems are reactive, not proactive, as they are directed at detecting misinformation once it is published and not in anticipation of its appearance. All these factors highlight the importance of the need to conduct future research to incorporate the concepts of transparency, flexibility, and cross-linguistic soundness into the framework of fake news detection so that it is not only effective but also reliable in practical use.

2.9. Summary and Research Gaps

The review of fake news detection with machine learning shows impressive development in the last ten years, where the focus shifted on the basic classification of lexical entities to the advanced, multi-view, and multi-modal deep learning networks. Initial studies were devoted to extraction of textual features with classical machine learning algorithms SVM, Naive Bayes and

Passive-Aggressive Classifiers [7]. These techniques were based on manual based linguistic indications and stylistic indicators and they though interpretable, were not flexible to invisible information. The next research trend adopted deep learning, which allowed automatic representation learning and semantic reasoning with models such as CNNs, LSTMs, and transformers [5], [9]. This change had a significant positive impact on the accuracy of detection and its robustness, but also created problems in the explainability and data dependency. Multi-view and graph-based studies, i.e., a study by Pereira et al. [1], a study by Bazmi and Shakery [4], have redefined the fake news detection as a relational issue. These models combine textual message, user action, and network, proving that it is impossible to comprehend misinformation out of its social and temporal context. In the meantime, the unsuitability of single-size-is-all models is emphasized by domain-specific studies (e.g., Alkudah et al. [6]). The language peculiarities, differences in culture, and inaccessibility of data in non-English language remain obstacles to generalization. Likewise, researchers of multimedia misinformation [3] indicate the increasing sophistication of fake news in question whereby manipulation is conducted in the forms of videos, audio, and images instead of text only.

In spite of these developments, there are still a number of research gaps. First, cross-lingual transferability is restricted by the absence of big, balanced and multi-lingual datasets. The majority of the currently available corpora are still English-oriented, hindering further development of the areas, where misinformation factors are linguistically specific. Second, hybrid and deep models have high accuracy but lower interpretability, which makes automated systems less transparent and trustworthy [8]. Third, the detection of fake news is mainly reactive, and it is more oriented towards classification instead of early warning or prevention systems. Fourth, adversarial robustness has not been

thoroughly studied so far fake content creators are becoming more and more able to take advantage of model vulnerabilities using generative AI methods [10].

Another multimodal fake news detector is proposed by Lu and Yao and combines multi-head attention mechanisms and residual convolutional networks. They take text, images, and video as their input and allow cross-modal alignment and further semantic fusion. The model shows good results in the datasets of LIAR (0.977 accuracy, 0.986 precision, 0.969 recall, F1 0.924), FakeNewsNet (0.986 accuracy, 0.979 precision, 0.965 recall, F1 0.900), and Weibo (accuracy 0.951, precision 0.961, recall 0.926, F1 0.843). It performs better than strong baselines like It has been shown that the study is exceptionally robust, resistant to perturbation of the data, and remarkably fast in inference (less than 0.02 s). The authors conclude that a combination of multimodal fusion, residual architectures, and attention can be used to produce a more reliable and deployable fake news detection system in a variety of media settings [9].

Al-Obaidi and Çağlıkantar compare a wide variety of machine learning and deep learning-based fake news detectors that are based on ISOT, FakeNewsNet, and FakeNewsNet + extensions (Half-True and Barely-True). Their work focuses on binary and multiclass detection, SVM, RF, XGBoost, LSTM, CNN-LSTM, and CNN-GRU are the testing algorithms. Findings indicate extremely high performance of LSTM-based models with an accuracy of 99.95% (ISOT) and 99.97% (FakeNewsNet). The authors point out that the performance differs greatly in different datasets, and the incorporation of nuanced classes (Half-True, Barely-True) makes the models more complex and lowers their accuracy, especially in the case of LSTM AutoEncoder. Their research shows how the dataset structure, the ability to model text only, and alignment between models and datasets play a crucial role in the achievement of the best fake news detection accuracy [10].

To sum up, in all studies reviewed, the conclusion is that there is a necessity in explainable, adaptable and culturally inclusive detection frameworks. The research approach to be undertaken in future should revolve around cross-domain generalization, low-resource language modeling, and ethical AI incorporation as a way of improving the technical effectiveness and societal credibility of fake news detection models.

Table 1 Summary of Key Studies in Fake News Detection Using Machine Learning

Ref.	Author(s) & Year	Dataset Used	Algorithm / Technique	Evaluation Results	Main Conclusion
[1]	Pereira et al. (2025)	FakeNewsNet & social graphs	Multi-view autoencoder (MV-AE)	Accuracy 93%	Combining content & social signals in shared latent space improves detection.
[2]	Ying et al. (2023)	PolitiFact, GossipCop	Bootstrapped multi-view representation	90% accuracy	Aligning multiple feature views reduces noise and boosts generalization.
[3]	Min et al. (2022)	Twitter & Weibo data	Post-user interaction network (PUI-Net)	$F1 \approx 0.89$	User interaction structure offers high predictive value beyond text content.
[4]	Bazmi & Shakery (2023)	Credibility-labeled corpus	Multi-view co-attention network	Accuracy 91%	Topic-specific modeling of users and sources enhances relational credibility scoring.
[5]	Zhang et al. (2024)	FakeNewsNet	GBCA: GCN + BERT + co-attention	Accuracy 95%	Integrating graph and textual features significantly improves detection accuracy.
[6]	Alkudah et al. (2025)	Arabic Fake News Dataset	CNN, LSTM, SVM comparison	DL: 92%, ML: 85%	Deep learning surpasses ML for morphologically rich Arabic text.
[7]	Kesarwani et al. (2020)	Twitter fake news	K-Nearest Neighbors (KNN)	84% accuracy	KNN effective for small datasets but limited scalability.
[8]	Domenico et al. (2021)	Marketing misinformation datasets	Systematic literature review	—	Fake news thrives via emotional contagion and algorithmic amplification.
[9]	Lu, Y., & Yao, N. (2025)	•LIAR dataset (political short text)	• Multimodal residual	Best-case performance: • Accuracy: 0.977	Integrating residual CNNs with multimodal attention significantly

		• FakeNewsNet (English multimodal news) • Weibo dataset (Chinese social media multimodal data)	convolutional network • Multi-head attention mechanisms • Multimodal feature fusion (text + image + video) • Cross-modal alignment module	• Precision: 0.986 • Recall: 0.969 • F1 score: 0.924 Also shows strong robustness, stability, and fast response (<0.02 s).	improves fake news detection across text, image, and video. Model is robust, fast, and performs better than BERT, RoBERTa, XLNet, ERNIE, and GPT-3.5 baselines.
[10]	Al-Obaidi, S. A., & Çağlıkantar, T. (2024)	ISOT, FakeNewsNet, Dataset1, Dataset2 (ISOT + LIAR-PLUS: Half-True, Barely True)	Multiple ML models (SVM, LR, DT, RF, XGBoost, Gradient Boosting) + Deep Learning models (CNN, CNN-LSTM, CNN-GRU, LSTM, LSTM AutoEncoder)	• ISOT: LSTM accuracy 99.95% • FakeNewsNet: CNN-GRU accuracy 99.97% • Dataset1: CNN-LSTM & LSTM nearly 100% • Dataset2: LSTM 98.64% • LSTM AutoEncoder lowest performance (49%–62%)	Model performance depends heavily on dataset structure. LSTM and CNN-based hybrids outperform others, while LSTM AutoEncoder is less suitable for fake news detection. Multiclass extensions (Half-True, Barely True) improve nuance in prediction.

2.10. Chapter Summary

The conceptual, methodological, and empirical development of detecting fake news based on machine learning provided in the literature review in this chapter demonstrated a shift in simple text-based detectors to more advanced multi-view and multi-modal detectors that are based on deep learning. In all the reviewed papers, it becomes clear that the issue of misinformation is not limited to the linguistic manipulation but goes far deeper, into the social, contextual, and behavioral aspects. The chapter disclosed the way in which fake news has become a multidimensional research area involving the combination of natural language processing (NLP), network analysis, and modeling of cognition.

Initial works, e.g., by Kesarwani et al. [7], defined what feature-based learning would be based on, i.e., textual indicators of word frequency, sentiment, and syntax to learn on a conventional machine learning structure (SVM, KNN, Naive

Bayes). These models proved to be reasonable but had restrictions in the form of manual feature engineering and in-scaling. Later studies took the direction of deep learning, such as Zhang et al. [6] who used CNNs, LSTMs and Transformer-based models (e.g. BERT, Falcon) that can learn complex semantic patterns independently. The excellent results of these models, which in many situations reached the 90 percent mark, justified the shift to data-driven representation learning.

The next significant concept that was covered is the emergence of multi-view and graph-based models where papers like Pereira et al. [1], and Bazmi and Shakery [4] used a combination of both textual and social cues to formulate relational dependencies among users, sources and content. These models were more generalizable and interpretable because they thought of misinformation as being a network phenomenon and not an unchanging document classification problem. Also, domain and multilingual studies, such as those, by Alkudah et al. [6] has highlighted the linguistic and cultural peculiarity of misinformation, especially in the Arabic and other non-English languages, and have shown that there is a severe deficit in cross-lingual generalization.

Chapter 3: Research Methodology

3.1.Introduction

The chapter outline is the methodology of designing and testing fake news detection models that rely on machine learning (ML) and deep learning (DL) models. The aim of the study is to come up with an automated model that can identify both real and fake news articles using the ISOT Fake News Dataset. The approach is divided into several steps such as acquiring the dataset, data preprocessing, feature representation, model training and performance evaluation. All the stages are designed in order to be able to guarantee the reproducibility, scalability and reliability of the results.

The use of both ML and DL approaches is intended to deliver a comparative insight into the traditional models of features engineering and the recent neural model that is guided by learning the representations of the data. Machine learning methods like the Support Vector Machines (SVM), Logistic Regression, and Random Forest are chosen because they are interpretable and computationally efficient, whereas deep learning methods like the Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) are chosen due to their capability to learn semantic and contextual relationships in textual data. Through the investigation of the two paradigms, the study aims to establish the degree to which the automated extraction of features involving the use of deep networks can perform better than a manually designed method and still be explainable.

3.2. Dataset Description

Isot Fake News Database The main dataset that will be used in the current study and will be a strong benchmark in the evaluation of fake news detectors. This data set is comprised of the news items found in the credible and untrustworthy

internet sources over the period of 2016 and 2017. The real news subset has been created based on the Reuters.com, the internationally known and credible news agency whereas the fake news subset was found on the different web sites which are classified as misinformation sources by the Politifact and Wikipedia websites. The data is separated into two CSV files, True.csv which includes genuine news and Fake.csv which includes fake articles. Every file contains four important features namely, the title of the article, the main text, the category of the subject and the date when it was published and a combination of these features forms the complete text of the article, which can be used in the classification processes.

Overall, the collection that was used consists of about 44,898 news articles, which consist of 21,417 true and 23,481 fake entries. Under real-news category, there are 10,145 articles that fall under World News category and 11,272 articles that fall under Politics News category. The fake-news category, in its turn, contains 1,570 Government News, 778 Middle-East, 783 U.S. News, 4,459 Left-News, 6,841 Politics and 9,050 general News articles. This allocation proves that political and world news occupies the leading place in both categories indicating the thematic bias that is frequently inherent in the process of misinformation.

Each of the articles has four main features namely the title, text, subject, and date of publication. The information was pre-cleaned with the original style of writing and punctuation being retained to maintain authenticity. Table 2 presents the categories of elements of the dataset in a detailed breakdown.

Table 2 breakdown of the dataset categories

News Type	Total Articles	Category	Articles per Category
Real News	21,417	World News	10,145
		Politics News	11,272
Fake News	23,481	Government News	1,570

		Middle East	778
		U.S. News	783
		Left News	4,459
		Politics	6,841
		General News	9,050

The dataset provides a balanced and diverse set of topics, with politics and global news dominating both classes. The total sample size of 44,898 articles ensures sufficient variability for robust training and testing across algorithms.

3.3. Data Preprocessing

Preprocessing of data is an important stage of work on text classification because the quality and format of input data directly affect the accuracy of predictive models. The ISOT dataset is rather clean, but it is necessary to process it further to convert raw textual data into a computer-readable form that can be used in ML and DL algorithms. Text normalization, which allows elimination of HTML tags, URLs, numbers, and non-alphanumeric characters so as to normalize the corpus, and retain punctuation that can have stylistic importance, is the first step. This is done to remove insignificant noise, and to concentrate the model on useful linguistic patterns.

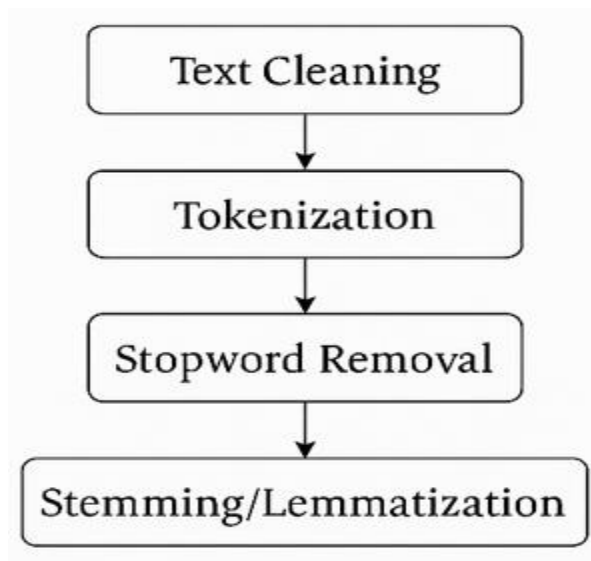


Figure 3 Preprocessing stages

The text is then normalized and tokenized to represent sentences as individual tokens (words) in order to be represented as vectors. The common stopwords are also omitted, including the, and, is, etc., because they do not add much to the classification. Stemming or lemmatization, based on the modeling technique, is used to reduce the words to their root or base form so that the various grammatical inflections of the same word are handled identically. There is an example of running, ran and runs which are all shortened to run. The step increases feature extraction consistency.

After being cleaned, the textual information is coded numerically. In machine learning models, the text is transformed into high-dimensional sparse matrices to obtain the relative importance of each term by using the Term Frequency-Inverse Document Frequency (TF-IDF) technique. In case of deep learning models, every token is then encoded into a high-dimensional medium-sized vector with pre-trained word embedding like Word2Vec or GloVe, and neural networks can learn meaningful word-word relationships.

Lastly, the dataset is split into training (80) and testing (20) subsets with the help of stratified sampling to ensure equal representation of both fake and real articles. This will be to guarantee an objective review of model performance. The preprocessing pipeline, by extension, ensures that every model is presented with an equal amount of structurally presented and semantically enriched input data, allowing the accurate identification of deceptive patterns. Overall, successful preprocessing can convert the ISOT corpus of raw human text into a structured information-rich format that is optimal to both statistical and neural analysis of information.

3.4. Model Development

The phase of model development will be concerned with designing and training machine learning and deep learning-based architectures to classify fake news.

The combination of these two paradigms makes it possible to perform a comparative analysis of the traditional models that are based on explicit feature engineering in contrast to neural networks that can learn to represent something automatically.

Three models are applied in the machine learning phase, including the Logistic Regression, Support Vector Machine (SVM), and Random Forest. Logistic Regression is a linear base on which probabilities are estimated of being a member of a certain class whereas SVM maximizes the distance between fake and real news by separating them with the help of a kernel. Ensemble method, namely, the Random Forest, is a non-linear correlating method that builds many decision trees together. The training of all ML models is done by using TF-IDF feature matrices and optimizing the grid search with 5-fold cross-validation to narrow down on hyper parameters like regularization strength, type of kernel and depth of the tree.

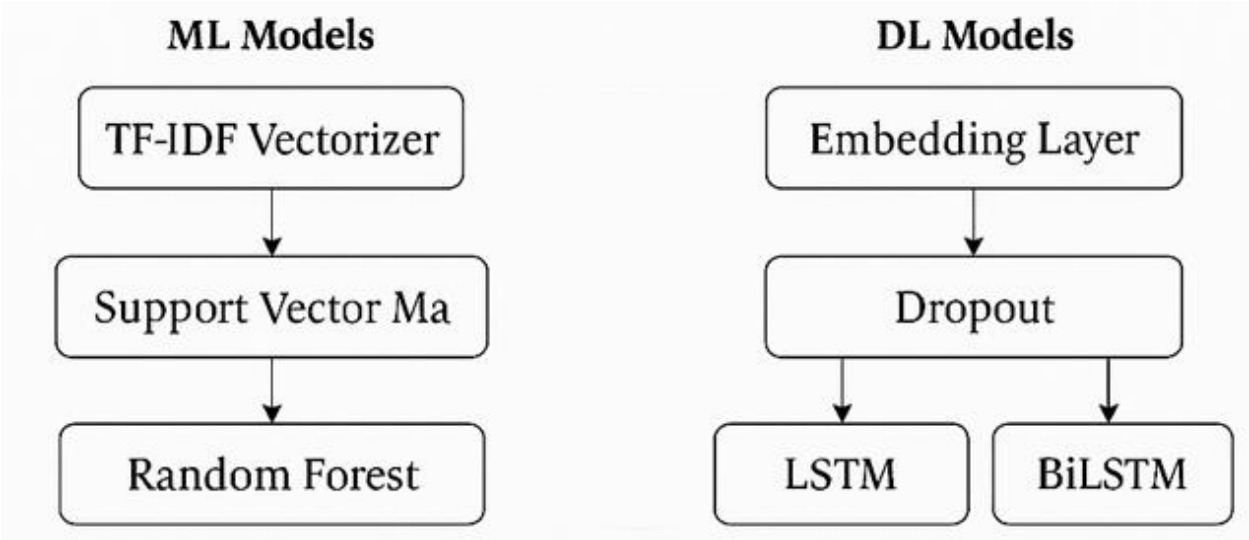


Figure 4 Model Architecture

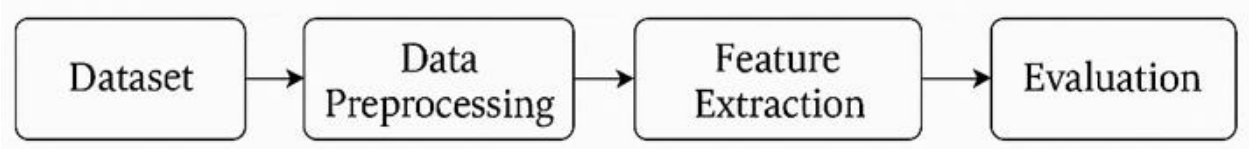


Figure 5 Work Flow

Deep learning stage uses successive neural structures that have the ability to learn the connection of context of the word embeddings. The LSTM model is chosen because of its potential to learn long term dependencies in text and Bidirectional LSTM (BiLSTM) improves the potential of LSTM because it allows input sequences to be processed in both forward and backward directions to gain deeper insight into their semantics. Also, a 1D Convolutional Neural Network (CNN) is employed to obtain local textual patterns like n-grams that can be helpful in detecting stylistic and syntactic clues. Both models have layers of embedding word vectors that are pretrained, dropout layers to regularize the model, and the Adam optimizer to optimize gradient descent.

The Multi-epoch training is done with mini-batches of 64 samples per epoch and epochs of 10-15. Models are all tested on the same test set. This two-framework design gives an idea of the comparison between the feature-based ML algorithms and the context-aware neural networks in identifying misleading content. Through a comparison of their merits and demerits, this study determines a holistic concept of the way computational models decipher linguistic deception in the digital media.

3.5. Evaluation Metrics

To determine the effectiveness and reliability of the developed models several evaluation metrics are used, which guarantees not only quantitative rigor but also clarity of interpretation. Accuracy is the main concept that will be evaluated, as it represents the general ratio of accurately defined cases. Nevertheless, due to the fact that accuracy might not be indicative of the performance in terms of classes, other measures including precision, recall, and F1-score are added. Precision determines the number of predicted fake articles that are actually fake and recall is a metric used to determine how effective the model is at identifying all fake news in the data. The harmonic mean of precision and recall is also

called F1-score which gives an unbiased perspective and is useful especially in datasets where there can be an imbalance in class.

Each model is put into a confusion matrix to visualize the true positives, false positives, true negatives as well as false negatives in order to give a detailed picture of the classification behavior. Cross-validation is also used to evaluate machine learning models in order to determine the performance of these models in unseen data. In the case of deep learning models, loss curves of training and validation are tracked on an epoch-by-epoch basis in order to identify overfitting or underfitting behaviors.

ML and DL models are then compared in terms of performance. Deep learning models are anticipated to have higher accuracy and F1-scores given that they have the ability to learn representations. Nevertheless, ML models, including SVM and Random Forest, can prove to be more interpretable and efficient in their computations. The assessment model thus puts emphasis on predictive capability as well as explainability and is in accordance with the aim of the study to determine the best but understandable fake news detection method.

The standard measures that are used to measure the performance of the models are as follows:

- Accuracy (ACC)- percentage of correctly predicted samples.
- Precision (P) - percentage of the fake news that is predicted to be fake.
- Recall (R) - percentage of fake news that was accurately recognized.
- F1-Score (F1)- harmonic mean between precision and recall.
- Confusion Matrix - to visualize the truth predictions and false predictions.

The comparison between ML and DL models is made in terms of these metrics to decide which method has a higher generalization and more robustness.

3.6.Research Roadmap

Table 3 Research Roadmap

Week	Focus Area	Detailed Tasks	Expected Outcomes
Week 1	Literature Collection & Organization	- Collect key papers on fake news detection (ML/DL, transformers, XAI, multimodal). - Gather theoretical sources on misinformation, linguistics, psychology. - Build a structured library (Zotero/Mendeley).	A complete database of 20–30 high-quality academic papers.
Week 2	Literature Review, Synthesis & Gap Analysis	- Break down literature into themes (content-based, social-based, multimodal, XAI). - Summarize methods, datasets, metrics, limitations. - Identify research gaps directly linked to your model choices. - Finalize the theoretical framework section.	Completed literature review + table of research gaps + clear justification for your methodology.
Week 3	Dataset Expansion & Preprocessing	- Collect additional datasets (FakeNewsNet, LIAR, multilingual). - Clean, normalize, tokenize text. - Create domain-aware train/validation/test splits. - Initial data exploration.	Clean merged dataset ready for model training.
Week 4	Model Setup (ML + LSTM/BiLSTM + Transformers)	- Configure baseline ML models (LR, SVM, Random Forest). - Prepare LSTM/BiLSTM architectures. - Set up BERT/RoBERTa/DeBERTa transformer models. - Build unified training pipeline.	All models (classical + deep + transformer) ready for training.
Week 5	Model Training (Phase 1 — Baselines)	- Train ML models on ISOT dataset. - Train LSTM/BiLSTM architectures. - Tune hyperparameters. - Save baseline performance metrics.	Baseline performance benchmark established.
Week 6	Model Training (Phase 2 — Transformers)	- Train transformer models on ISOT + expanded datasets. - Optimize training and evaluate using F1, AUC, accuracy. - Compare against baseline ML/DL models.	High-performance transformer results + comparison table.
Week 7	Explainability & Interpretability (XAI)	- Apply SHAP, LIME, Integrated Gradients. - Add attention visualization for BERT models. - Generate explanation plots and word-importance maps. - Document findings for the thesis.	Complete explainability section + visual evidence for thesis.

Week 8	Adversarial Robustness & Stress Testing	• Create adversarial examples (paraphrasing, synonym replacement, style attacks). • Evaluate model vulnerability. • Apply adversarial training to improve robustness.	Robustness report + improved adversarial-resistant model.
Week 9	Evaluation, Benchmarking & Error Analysis	• Cross-dataset evaluation (generalization). • Confusion matrices, ROC curves, precision/recall tables. • Perform ablation studies. • Analyze model misclassifications.	Final experimental results + comprehensive evaluation chapter.
Week 10	Documentation, Results Writing & Publication Prep	• Write final methodology, results, discussion, and future-work sections. • Prepare thesis tables, figures, charts. • Draft a journal/conference paper summarizing contributions.	Fully written thesis chapters + publish-ready research paper.

3.7. Summary

This chapter explained how the fake news would be detected using the ISOT dataset. The study will combine old-fashioned ML algorithms with the new DL models to determine which of them provides the most efficient and understandable results. The hybrid approach design makes sure that the experimentation is complete not only in terms of feature-engineered but also representation-learning paradigms. Chapter four will provide the analysis and results of the experiment, compare the algorithmic performance and commenting on the implications of the findings as they relate to the problem of misinformation detection.

References

1. Ingryd V. S. T. Pereira, George D. C. Cavalcanti, Rafael M. O. Cruz, Multi-view autoencoders for Fake News Detection, Preprint · April 2025, research gate DOI: 10.48550/arXiv.2504.08102
2. Q. Ying, X. Hu, Y. Zhou, Z. Qian, D. Zeng, and S. Ge, “Bootstrapping multi-view representations for fake news detection,” in AAAI Conference on Artificial Intelligence, vol. 37, no. 4, 2023, pp. 5384–5392.
3. Min E, Rong Y, Bian Y, Xu T, Zhao P, Huang J, Ananiadou S (2022) Divide-and-conquer: post-user interaction network for fake news detection on social media. In: Proceedings of the ACM Web Conference 2022, pp. 1148–1158
4. Bazmi, M. Asadpour, and A. Shakery, “Multi-view co-attention network for fake news detection by modeling topic-specific user and news source credibility,” Information Processing & Management, vol. 60, no. 1, p. 103146, 2023.
5. Zhang Z, Lv Q, Jia X, Yun W, Miao G, Mao Z, Wu G (2024) Gbca: graph convolution network and bert combined with co-attention for fake news detection. Pattern Recogn Lett 180:26–32
6. Alkudah NM, Idris NB, Abushariah MAM. 2025. Fake news detection in Arabic media: comparative analysis of machine learning and deep learning algorithms using the Arabic Fake News Dataset. PeerJ Comput. Sci. 11:e3272 DOI 10.7717/peerjcs.3272.
7. Kesarwani, S. S. Chauhan and A. R. Nair, "Fake News Detection on Social Media using K- Nearest Neighbor Classifier," 2020 International Conference on Advances in Computing and Communication Engineering (ICACCE), 2020, pp. 1-4, 10.1109/ICACCE49060.2020.9154997
8. Domenico GD, Sit J, Ishizaka A, Nunan D. 2021. Fake news, social media and marketing: a systematic review. Journal of Business Research 124(2):329–341 DOI 10.1016/j.jbusres.2020.11.037.
9. Lu, Y., & Yao, N. (2025). *A fake news detection model using the integration of multimodal attention mechanism and residual convolutional network*. *Scientific Reports*, 15, 20544. <https://doi.org/10.1038/s41598-025-05702-w>

10. Al-Obaidi, S. A., & Çağlıkantar, T. (2024). *Automated fake news detection system*. *Iraqi Journal for Computer Science and Mathematics*, 5(4), 12–26.
<https://doi.org/10.52866/2788-7421.1200>