REPUBLIC OF TÜRKİYE

ALTINBAŞ UNIVERSITY

Institute of Graduate Studies

Electrical and Computer Engineering

MFF-LSTM: DESIGNING A MULTI-SCALE FEATURE FUSION-BASED LONG SHORT-TERM MEMORY WITH DIVERGENT FEATURES FOR FAKE NEWS DETECTION SYSTEM

Mustafa Saeb Sedeeq ALSAFAWI

Master’s Thesis

Supervisor

Asst. Prof. Dr. Oğuz KARAN

İstanbul, 2023

MFF-LSTM: DESIGNING A MULTI-SCALE FEATURE FUSION-BASED LONG SHORT-TERM MEMORY WITH DIVERGENT FEATURES FOR FAKE NEWS DETECTION SYSTEM

Mustafa Saeb Sedeeq ALSAFAWI

Electrical and Computer Engineering

Master's Thesis

ALTINBAŞ UNIVERSITY

2023

The thesis titled MFF-LSTM: DESIGNING A MULTI-SCALE FEATURE FUSION-BASED LONG SHORT-TERM MEMORY WITH DIVERGENT FEATURES FOR FAKE NEWS DETECTION SYSTEM Prepared by Mustafa Saeb Sedeeq ALSAFAWI and submitted on 29/12/2023 has been **accepted unanimously** for the degree of Master of Science/PhD in Electrical and Computer Engineering

Asst. Prof. Dr. Oğuz KARAN

Supervisor

Thesis Defense Committee Members:

Asst. Prof. Dr. Oğuz KARAN	Department of Software Engineering, Altınbaş University	_____
Assoc. Prof. Dr. Sefer KURNAZ	Department of Computer Engineering, Altınbaş University	_____
Asst. Prof. Dr. Zeynep ALTAN	Department of Software Engineering, Beykent University	_____

I hereby declare that this thesis meets all format and submission requirements of a Master's thesis.

Submission date of the thesis to Institute of Graduate Studies: ___/___/___

I hereby declare that all information/data presented in this graduation project has been obtained in full accordance with academic rules and ethical conduct. I also declare all unoriginal materials and conclusions have been cited in the text and all references mentioned in the Reference List have been cited in the text, and vice versa as required by the abovementioned rules and conduct.

Mustafa Saeb Sedeeq ALSAFAWI

Signature

DEDICATION

I extend my thanks and gratitude to my supervisor, Dr. Oğuz KARAN, for agreeing to supervise my thesis duringmy master’s studies and providingme with full support. I also dedicate this thesis to my family for standing by me and encouraging me to obtain a master's degree. I also thank all my friends and colleagues who supported me.

ABSTRACT

MFF-LSTM: DESIGNING A MULTI-SCALE FEATURE FUSION-BASED LONG SHORT-TERM MEMORY WITH DIVERGENT FEATURES FOR FAKE NEWS DETECTION SYSTEM

ALSAFAWI, Mustafa Saeb Sedeeq

M.Sc., Electrical and Computer Engineering, Altınbaş University,

Supervisor: Asst. Prof. Dr. Oğuz KARAN

Date: December /2023

Pages: 68

The rise in the usage of media has led to a significant increase in the raise of false information, making it imperative to combat this issue and reduce our reliance on such unreliable sources. Fake news can mislead people, spread rumors, and even impact the positions of political leaders. Detecting fake news has become crucial in this digital era, with direct messaging platforms and social media playing a major role in its proliferation. Various innovative techniques have been suggested to determine fake news, making the endeavor both intriguing and challenging. Hence, this synopsis aims to develop the adaptive learning model with multiscale feature fusion for fake news detection. The proposed system constitutes “text collection, text pre-processing, feature extraction and detection”. Initially, the text input is collected from the benchmark datasets, which is then followed by the text stage of pre-processing. Here, the pre-processed text is obtained that is fed into the feature extraction phases. The three feature extraction techniques such as “Bidirectional Encoder Representations from Transformers (BERT), Term Frequency-Inverse Document Frequency (TF-IDF) and GloVe Embedding” are employed to provide the feature set 1, 2 and 3. Finally, these resultant features are given to “Multiscale Feature Fusion based Long Short Term Memory (MFF-LSTM)”, where the features are fused together in multiscale manner and detection is taken place by LSTM. Therefore, the system evaluation is done by considering the distinct measures and compared among traditional approaches. Hence, the recommended model attains the desired results to detect the fake news that helps to evade the exploration of any false information.

Keywords: Fake News, Detection System , MFF, LSTM, Multi-scale Feature

TABLE OF CONTENTS

Pages

ABSTRACT ..... **vi**

LIST OF FIGURES ..... **x**

ABBREVIATIONS ..... **xii**

LIST OF SYMBOLS **1**

1. INTRODUCTION **2**

1.1 PROBLEM STATEMENT 3

1.2 RESEARCH MOTIVATION 4

1.3 RESEARCH OBJECTIVES 7

1.4 CONTRIBUTION 7

1.5 ISSUES REGARDS TO DEFINE THE FAKE NEWS 8

1.6 ISSUES REGARDS TO DETECT THE FAKE NEWS 10

2. LITERATURE REVIEW **12**

3. MATERIAL AND METHOD **17**

3.1 DESIGNING OF FAKE NEWS DETECTION MODEL USING MULTI SCALE FEATURE FUSION-BASED DEEP LEARNING METHODOLOGY 17

3.1.1 Experimented Dataset Details 17

3.1.2 Brief Description of Proposed System 18

3.1.3 Text Pre-processing 21

3.1.3.1 Punctuation and special character removal 21

3.1.3.2 Remove redundant and inappropriate data 22

3.1.3.3 Stop words removal 22

3.1.3.4 Stemming 23

3.2 EXTRACTING FEATURE TECHNIQUES OF BERT, TF-IDF AND GLOVE EMBEDDING TO DETECT THE FAKE NEWS 24

3.2.1 Extracting Feature Techniques of BERT, TF-IDF and Glove Embedding to Detect the Fake News 24

3.2.1.1 BERT-based features 24

3.2.1.2 TF-IDF-based features 26

3.2.1.3 GloVe-based features 28

3.2.1.4 GloVe-based features 30

3.3 FAKE NEWS DETECTION BY APPLYING THE MULTI-SCALE FEATURE FUSION-AIDED LONG SHORT-TERM MEMORY 33

3.3.1 Multi-scale Concept and its Benefits 33

3.3.2 Basic LSTM 34

4. RESULTS AND DISCUSSION 38

4.1 PERFORMANCE METRICS 38

4.2 ROC ANALYSIS FOR DATASET 1 AND DATASET 2 39

4.3 COMPARATIVE ANALYSIS OF BATCH SIZE OVER TRADITIONAL CLASSIFIER IN DATASET 1 AND DATASET 2 40

4.4 PERFORMANCE ANALYSIS ON FEATURE EXTRACTION FOR DATASET 1 AND DATASET 2 45

4.5 COMPARATIVE ANALYSIS OF LEARNING PERCENTAGE OVER TRADITIONAL CLASSIFIER IN DATASET 1 AND DATASET 2 48

4.5 OVERALL PERFORMANCE ANALYSIS OF THE PROPOSED FAKE NEWS DETECTION MODEL FOR DATASET 1 AND DATASET 2 52

5. CONCLUSION 54

REFERENCES 56

LIST OF TABLES

Pages

Table 2.1: Characteristics and Defects of Existing Fake News Detection Model using Deep Learning. 16

Table 4.1: Overall Performance Analysis of the Proposed Fake News Detection Model for Dataset. 53

Table 4.2: Overall Performance Analysis of the Proposed Fake News Detection Model for Dataset. 53

LIST OF FIGURES

Pages

Figure 3.1 The Architecture View of The Suggested Detection Model of Fake News Based on Deep Learning. 19

Figure 3.2: Diagrammatic View of BERT Based Feature. 26

Figure 3.3: Diagrammatic View of TF-IDF Based Feature. 28

Figure 3.4: Diagrammatic View of BERT Based Feature. 30

Figure 3.5: Diagrammatic View of GloVe Based Feature. 33

Figure 3.6: The Architecture View of LSTM. 36

Figure 3.7: A Proposed View of MFF-LSTM For Fake News Detection. 37

Figure 4.1: The ROC Estimation of the Developed Fake News Detection Model Over Numerous Traditional Classifiers Regarding "(a)Dataset 1 and (b) Dataset 2". 40

Figure 4.2: The Batch Size Evaluation of the Suggested Fake News Detection Model Over Diverse Conventional Classifier for Dataset 1 Concerning " (a) Accuracy, (b) F1-score, (c) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NPV, (h) Precision, (i) Sensitivity". 41

Figure 4.3: The Batch Size Evaluation of the Suggested Fake News Detection Model over Diverse Traditional Classifier for Cataset 2 Concerning " (a) Accuracy, (b) F1-score, (c) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NPV, (h) Precision, (i) Sensitivity. 43

Figure 4.4: Figure 10:Comparative Evaluation of Recommended Fake News Detection Model's Feature Extraction over Diverse Conventional Classifier for Dataset 1 Concerning " (a) Accuracy, (b) F1-score, (c) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NPV, (h) Precision, (i). 45

Figure 4.5: Comparative Evaluation of the Suggested Fake News Detection Model's Feature Extraction over Diverse Traditional Classifier for Dataset 2 Concerning " (a) Accuracy, (b) F1-score, (c) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NPV, (h) Precision, (i) Sensitiv. 47

Figure 4.6: The Learning Percentage Revaluation of the Suggested Fake News Detection Model's Feature Extraction over Diverse Traditional Classifier for Dataset 1 Concerning " (a) Accuracy, (b) F1-score, (c) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NPV, (h) Precision, (i) Sensitivity". 49

Figure 4.7: The Learning Percentage Evaluation of The Suggested Fake News Detection Model's Feature Extraction Over Diverse Traditional Classifier for Dataset 2 Concerning " (a) Accuracy, (b) F1-Score, (c) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NPV, (h) Precision, (i) Sensitivity". 51

ABBREVIATIONS

LR	:	Logistic Regression
SVM	:	Support Vector Machine
KNN	:	K-Nearest Neighbor
DT	:	Decision Tree
RF	:	Random Forest
LSTM	:	Long Short-Term Memory
GRU	:	Gated Recurrent Network

LIST OF SYMBOLS

μ : Electron Mobility

λ : Wavelength

ω : Angular Frequency

1. INTRODUCTION

In the era of advanced technology, an overwhelming volume of data is generated online daily. The rapid improvements in technology for communication and mobile internet access have made the impact of social media's widespread influence an essential part of our everyday life [9]. Social networking sites on the internet have become the world's favorite news sources, particularly for those in underdeveloped nations [10]. Consequently, these platforms have provided a global stage where individuals from any corner of the world can leverage social media networks to disseminate statements and propagate fake news through various networking sites. This phenomenon is often exploited for diverse and sometimes illicit purposes [11]. These days, inaccurate data is widely disseminated due to users' unfettered freedom to produce content on internet platforms, such as social networking sites and news sites. Readers have been misled by intentional attempts to disseminate incorrect information [12]. Academics and business alike are becoming more worried about the spread of fake news. Human verification is one proposed remedy for this issue [13]. It is commonly acknowledged that fake news poses a major risk to journalism, the rule of law, and international trade while also generating significant collateral harm. The risk of producing and disseminating false information has increased due to the rising popularity of social networking platforms [14]. Social media networks allow users to openly express their perspectives and views, frequently without scrutiny. Interestingly, certain news reports shared on platforms like Twitter and Facebook garner more views than those published directly on official news websites [15].

The openness of social media, coupled with its vast user base and the intricate network of sources, has facilitated the proliferation of numerous fake news stories [16]. Unscrupulous individuals exploit these widespread false narratives to deceive the public, posing significant risks to society and potentially leading to substantial economic losses [17]. Fake news is widely acknowledged as a major global threat, impacting commerce, journalism, and democracy, resulting in significant collateral damage [18]. Furthermore, the nature of real-time fake news on social media adds to the challenge of identifying and combating online misinformation. The effectiveness of expert fact-checking is often limited due to its low efficiency. It is crucial to create effective and trustworthy false news detection mechanisms because of the negative impacts fake news has on society [19]. The time and knowledge needed by average users to confirm the legitimacy of each item of info they see is lacking

[20]. Therefore, in order to guarantee that users obtain correct and true information, it is imperative to identify bogus news on online services. In present times, advanced DNN have emerged as a solution to this challenge. These networks eliminate the necessity for manually crafting features, as they can automatically capture intricate linguistic patterns, offering a more effective approach to identifying fake news [21].

Various classification approaches, including “Logistic Regression (LR), Support Vector Machine (SVM), k-Nearest Neighbor (k-NN), Decision Tree (DT), Random Forest (RF), LSTM and Gated Recurrent Network (GRU) have been explored for fake news detection [22]. The majority of fake news detection tools use machine learning approaches to help consumers filter information and assess the veracity of news stories [23]. In addition, in order to extract sequence features needed for identifying the appearance of bogus news, researchers have dug into RNN and GRU. Some studies have also introduced CNNs to capture high-level representations from social media posts for identifying fake news [24].However, these existing models often exhibit inconsistent performance across datasets of varying sizes and domains, making the task of evaluating news authenticity complex, even for automated systems [25]. To address this limitation, suggest a deep learning model for fake news detection. This helps to improve the performance of detection.

1.1 PROBLEM STATEMENT

In the current generation, the social media popularity has increased and the people changed the way to access the news. Most of the people reading the news in the online applications and there is the chance for misinformation and wrong news. The main issue is disinformation, which has drawn a lot of attention from the educational and business worlds. Unverified material is frequently disseminated on social media platforms with little regard for its veracity. The issue of identifying false news has gotten difficult, and a lot of recent study has concentrated on building models by taking abstract characteristics out of each post's content while ignoring the inherent semantic construction, which includes latent subjects and other concepts. Table I lists the characteristics and problems of the many approaches that have been used. CNN-RNN [1], is helpful for issues with regression and classification because it allows for both immediate and sequential properties of input data. However, it requires an extensive training set and has trouble determining the ideal hyper parameters for every problem and sample. Compared to other classification algorithms, random forest [2], has the best ratings for classification and is also more reliable and

accurate. Although it employs several decision tree structures and performs better with larger datasets, it is faster than models of classification. OPCNN-FAKE [3], extracts both a high-level and lower-level characteristics and performs better during cross-validation methods However, a large amount of knowledge for training is needed, and the quantity needed varies depending on how difficult the task is. PCA [4], provides more contextual features for detection and achieved better accuracy than another model. But it leads to loss of information and required data normalization. Four distinct tasks—bias identification, click bait identification, analysis of sentiment, and poison detection—are covered by representation learning [5]. However, its capability of representation is restricted. While MTMN [6], enhances the efficacy of fake news identification, extracting complete and supplementary data remains challenging. Large volumes of unlabelled data are precisely recognized using weakly supervised learning [7], which is then utilized for data identification. However, it solely takes into account machine learning methods for annotating data. While it can extract contextually-dependent features amongst words and attain greater accuracy on four raw data, BBC-FND [8], is not geographically stable to the input data. The study presented a deep learning technique to address the difficulties in detecting fake news in order to address the aforementioned issues.

1.2 RESEARCH MOTIVATION

The proliferation of widespread and digital information use of social media have transformed the way to consume news [41]. To counter this phenomenon, the development of fake news detectors has gained prominence. Some of the most important motivation and challenges are formulated as follows.

a. The primary motivation behind the creation of fake news detectors is to preserve the integrity of information. With the ease of content creation and distribution on the internet, distinguishing between genuine and fabricated news has become increasingly challenging. Fake news detectors aim to safeguard the accuracy and reliability of information, thereby upholding the public's trust in the media.
b. Fake news has profound societal consequences, ranging from influencing political opinions to exacerbating social tensions. The motivation to develop effective detectors stems from a desire to mitigate these impacts and prevent the manipulation of public

discourse. By identifying and flagging false information, these detectors contribute to the overall health of democratic societies.

c. Fake news detectors play a role in promoting media literacy by encouraging individuals to critically evaluate the information they encounter. The development of such tools serves as a reminder of the importance of verifying sources and understanding the context of news stories, empowering users to make more informed decisions about the information they consume.
d. Ensuring that information disseminated through news channels is accurate and reliable is crucial for maintaining the integrity of public discourse. The spread of fake news erodes public trust in media, institutions, and the democratic process, emphasizing the need for effective countermeasures.
e. Fake news can be weaponized to manipulate public opinion, influence elections, and sway decision-making processes, posing a threat to the democratic ideals of informed citizenry.
f. Fake news often thrives within ideological echo chambers, exacerbating societal divisions. Detecting and debunking false information helps counteract the reinforcement of polarized viewpoints.
g. False information can lead to real-world consequences, such as violence, discrimination, or panic [42]. Developing fake news detectors is motivated by the desire to minimize the potential harm caused by misinformation.
h. The development of fake news detectors goes hand-in-hand with efforts to enhance media literacy. By raising awareness about the existence of misinformation, people can become more discerning consumers of news.
i. Perpetrators of fake news continually adapt their tactics to circumvent detection. Staying ahead of these evolving strategies poses a significant challenge for fake news detectors. Machine learning algorithms must constantly evolve to identify new patterns and techniques employed by those disseminating false information.
j. Fake news detection requires a delicate balance between accuracy and speed. While it is essential to swiftly identify and flag misinformation, the risk of false positives can erode public trust if innocent content is wrongly labelled as fake. Striking the right balance between speed and accuracy remains an ongoing challenge in the development of effective detection systems.

k. The advent of deep fake technology and advanced manipulation techniques further complicates the detection of fake news. Detecting subtle alterations in audio, video, or even text requires sophisticated algorithms capable of discerning between authentic and manipulated content. Developing robust detectors that can handle these challenges is a pressing concern.
l. Fake news detectors must operate ethically and without bias to maintain public trust. Avoiding discrimination based on political or ideological affiliations while accurately identifying misinformation is a complex task. Striking a balance between neutrality and effective detection remains an ongoing challenge in the development and deployment of these tools.
m. Those spreading misinformation continually adapt their strategies, making it challenging for detectors to keep pace with the changing landscape of fake news.
n. Distinguishing between satirical content and actual misinformation can be challenging due to the nuanced nature of certain types of false information.
o. Developing algorithms to detect fake news involves risks of unintentional biases. Ensuring fairness and avoiding amplification of existing biases is a complex challenge.
p. The sheer volume and speed at which information is shared on social media platforms make it difficult to implement real-time detection systems without compromising accuracy.
q. While combating fake news, it is essential to respect privacy rights. Striking a balance between detecting misinformation and protecting user privacy is a delicate challenge.
r. Fake news often varies in its impact across different cultures and regions. Developing detectors that account for these variations is essential for their effectiveness on a global scale.
s. Even with effective detectors in place, getting users to adopt and trust these tools is a challenge. Educating users about the importance of fact-checking and critical thinking remains a significant hurdle.

Finally, it motivates to develop fake news detectors that arise from the need to protect the integrity of information and mitigate the societal impact of misinformation. However, addressing the evolving tactics of those spreading fake news and ensuring the ethical and unbiased operation of detection systems present significant challenges in this ongoing battle for information accuracy and reliability.

1.3 RESEARCH OBJECTIVES

The research objectives outlined for the suggested detection of fake news model provide a clear roadmap for the study. Let's break down each objective.

a. To obtain data from established benchmark datasets, likely containing examples of both real and fake news. The subsequent pre-processing tasks involve cleaning, formatting, and organizing the data to prepare it for analysis.
b. To investigate and apply a variety of deep learning feature extraction techniques. This step is crucial for capturing different facets of the text input, improving the model's capability to find patterns associated with fake news.
c. To implement deep learning based model to enhance the overall detection of fake news capability of the system.
d. To evaluate the proposed model by comparing its performance with traditional methods of fake news detection. This analysis aims to assess whether the developed model outperforms or offers advantages over conventional approaches in identifying and classifying fake news.

1.4 CONTRIBUTION

The proposed system outlined in the description appears to be a comprehensive approach for detecting fake news. The objectives of this system can be summarized as follows:

a. To create an adaptive detection of fake news model by using the deep learning that assists to easily and significantly identify the fake news to prevent the occurrence of misinformation. This model is expected to dynamically adjust and improve its performance based on the data it encounters.
b. To gather the input text from benchmark datasets that has been used to perform the pre-processing stages. This may involve tasks such as eliminating stop words, stemming, lemmatization, and handling special characters to prepare the text for further analysis.
c. To utilize three different feature extraction techniques that is BERT, TF-IDF and GloVe Embedding, where each of these techniques contributes a distinct set of features. This holistic approach enhances the model's ability to discern patterns and features relevant to fake news detection.
d. To propose a model of MFF-LSTM for fake news detection along with the concept of feature fusion. In the novel system, the features have been obtained via the different

scale of the network that is further combined and given for the subsequent layers. This integration of advanced techniques aims to improve the model's efficiency in discerning fake news from the input text.

e. To evaluate how well the suggested approach performs in comparison to other established methods for identifying false news. This stage is necessary to evaluate the produced system's quality and efficacy. The overall goal is to create a model with high accuracy for identifying fake news, which will aid in halting the rapid propagation of incorrect and inaccurate data.

1.5 ISSUES REGARDS TO DEFINE THE FAKE NEWS

Defining fake news is a complex task, as the term has evolved and taken on various meanings in different contexts. At its core, fake news refers to misinformation or disinformation presented as legitimate news [30]. The digital age has significantly amplified the challenges associated with defining fake news. The ease of creating and disseminating information online, coupled with the speed at which it can spread through social media, has made it a pervasive issue. Fake news can take many forms, ranging from fabricated stories and manipulated images to misleading headlines and distorted facts. Defining fake news presents a range of challenges and issues due to the dynamic nature of information dissemination, technological advancements, and the subjective interpretation of news. Here are some key issues regarding the definition of fake news.

a. Different individuals and groups may interpret the same news differently based on their perspectives, biases, and beliefs [39]. The interpretation of news is inherently subjective, influenced by personal beliefs, biases, and perspectives. This subjectivity adds a layer of complexity to the definition, making it challenging to establish universally agreed-upon criteria for discerning misinformation.

b. Technological innovations have not only facilitated the creation of content but have also made it increasingly sophisticated. Deep fakes, AI-generated text, and manipulated images contribute to the complexity of defining fake news. The integration of advanced technologies blurs the line between authentic and fabricated content, demanding a nuanced understanding of what constitutes misinformation.

c. The constantly changing landscape of information dissemination poses a significant challenge in defining fake news. The speed at which news travels in the digital age and the continuous updates to stories make it difficult to establish a static definition.

d. Political, cultural, and social contexts can influence how people perceive information, making it challenging to establish a universally agreed-upon definition.
e. The impact of fake news varies across different social and cultural contexts. Understanding the nuances of how information is interpreted and its societal impact is essential for crafting a comprehensive definition.
f. Distinguishing between intentionally deceptive information and unintentional errors or inaccuracies is difficult. While some instances of fake news are created with malicious intent, others may result from genuine mistakes, lack of verification, or editorial oversight.
g. The rapid evolution of technology and the media landscape has changed the way information is created, shared, and consumed. Traditional definitions of news may not adequately encompass the diverse forms of misinformation and disinformation present in the digital age.
h. Social media platforms play a significant role in the dissemination of information, and they often contribute to the rapid spread of fake news. Algorithms that prioritize engaging content may inadvertently promote sensational or misleading information.
i. Fake news is not confined to a specific region or country. Cultural differences and varying levels of media literacy across different societies make it difficult to formulate a single, universally applicable definition.
j. Those spreading misinformation continually adapt their tactics. As efforts to combat fake news evolve, so do the strategies employed by those seeking to deceive, making it challenging to create a static and comprehensive definition.
k. Disagreements over the veracity of certain information can arise due to differences in political ideologies, cultural backgrounds, and personal beliefs.

Addressing the issues related to fake news requires a nuanced understanding of the information landscape, media literacy, and effective fact-checking mechanisms. As society grapples with the consequences of misinformation, ongoing efforts are essential to refine and adapt the evaluation of fake news to meet the evolving challenges in the ever-changing media landscape.

1.6 ISSUES REGARDS TO DETECT THE FAKE NEWS

Detecting and combating fake news is a pressing challenge in the contemporary information landscape. As misinformation continues to proliferate, the ability to discern between authentic and deceptive content becomes increasingly crucial [40]. Various issues complicate the task of identifying fake news, ranging from the evolving tactics employed by purveyors of misinformation to the technological advancements that enable more sophisticated forms of deception [30]. This article explores key issues and challenges associated with detecting fake news, shedding light on the complexities that arise in the pursuit of fostering a more informed and resilient society.

a. Fake news manifests in various forms, including misleading headlines, fabricated content, altered images, and misinformation spread through social media. Detecting this diverse range of deceptive tactics requires versatile detection methods capable of addressing each manifestation effectively
b. The rapid dissemination of fake news through social media platforms necessitates timely detection. Algorithms must operate in real-time or near-real-time to keep pace with the speed at which misinformation can spread, preventing its potential impact on public perception.
c. Detecting fake news often requires an understanding of context, sarcasm, and linguistic nuances, which can be challenging for algorithms. Lack of context awareness may result in false positives or negatives, undermining the accuracy of detection systems.
d. Those creating fake news are often aware of detection methods and may actively try to evade them. Adversarial strategies, including intentionally adding misinformation to confuse algorithms, pose a constant challenge to accurate detection.
e. Balancing the need for effective detection with ethical considerations is crucial. Overly aggressive measures may infringe on freedom of speech, and false positives can harm legitimate content. Striking the right balance is essential to address misinformation without compromising democratic principles.
f. The training data used for machine learning models that contains biases, the models may inadvertently perpetuate and amplify those biases. Ensuring unbiased and diverse training datasets is critical to developing detection systems that are fair and effective across different demographics.

g. The sheer volume of information on the internet makes it challenging to monitor and analyse all content effectively. Efficient algorithms and scalable systems are necessary to sift through vast amounts of data in real-time to identify and counteract fake news.
h. Fake news can vary across cultures and regions, requiring detection systems to be sensitive to these variations to be effective globally. Understanding the cultural context in which misinformation operates enhances the accuracy and relevance of detection algorithms.
i. Algorithms must be capable of near-real-time or real-time analysis to keep up with the speed of information dissemination.
j. Detecting fake news often requires an understanding of context, sarcasm, and nuances in language, which can be challenging for algorithms. Lack of context awareness may lead to false positives or negatives. Those creating fake news are often aware of detection methods and may actively try to evade them.
k. Adversarial strategies, such as intentionally adding misinformation to confuse algorithms, pose a constant challenge. Balancing the need for effective detection with ethical considerations is crucial. Overly aggressive measures may infringe on freedom of speech, and false positives can harm legitimate content.
l. The sheer volume of information on the internet makes it challenging to monitor and analyze all content effectively. Fake news can vary across cultures and regions. Detection systems need to be sensitive to these variations to be effective globally.

Efforts to address these challenges often involve a combination of machine learning as well as NLP approach, and human moderation, this helps to improve the performance. Continuous research and collaboration between technology experts, researchers, and media organizations are essential to staying ahead of evolving fake news tactics.

2. LITERATURE REVIEW

In 2021, Abdul *et al.*, [1], have recommended a pioneering hybrid deep learning method, seamlessly integrated CNN and RNN, for the purpose of classified the fake news. This novel model underwent successful validation using two distinct fake news datasets. Notably, the achieved detection results surpassed those of traditional non-hybrid baseline methods, signifying a marked improvement in classification accuracy. Moreover, the model's performance was subjected to further scrutiny through experiments evaluating its generalization across diverse datasets. These additional tests yielded promising results, suggesting that the suggested hybrid deep learning model exhibited robust capabilities in adapting and performing well across various datasets. The positive outcomes of certain approaches in fake news detection emphasize their potential to bring about meaningful improvements in the effectiveness of detection methodologies and contribute to advancements in the broader field of identifying and classifying misinformation.

In 2021, Jiang *et al.*, [2], have evaluated the performance of machine learning method and three deep learning models using two distinct databases containing fake and real news. Employed hold-out cross-validation, utilized various text representation techniques, including TF-IDF, embedding methods and term frequency tailored for machine as well as deep learning models, correspondingly. To gauge the models' effectiveness, employed a comprehensive set of evaluation metrics, including precision, accuracy, F1-score and recall. Specifically, it achieved testing accuracies of 96.05% and 99.94% on the KDnugget and ISOT datasets, respectively. Notably, the efficiency of the suggested model surpassed that of baseline methods, underscored its efficacy in fake news detection. Based on these compelling findings, highly recommended the adoption of our novel stacking model for robust and accurate fake news detection applications.

In 2021, Saleh *et al.*, [3], have proposed a new approach integrating Deep learning as well as machine learning for the creation of a system that can detect disinformation with accuracy. A detailed comparison of OPCNN-FAKE's accuracy with six traditional machine learning approaches, RNN, and LSTM was conducted: “LR, DT, RF, KNN, Naive Bayes (NB) and SVM”. Four benchmark datasets dedicated to fake news were utilized for this comparative analysis. Optimization of model parameters was achieved through Hyperope techniques for ML, DL and Grid Search respectively. For normal machine learning, features were extracted using N-gram and TF-IDF, however for models based on deep learning, Glove word

embedding was used to encode the characteristics as a feature matrix. For verification of the findings, assessments of performance criteria including F1-measure, precision, accuracy and recall were used. The results showed that, on all datasets, the OPCNN-FAKE model repeatedly performed better than other algorithms. Interestingly, OPCNN-FAKE significantly outperformed other models in evaluation and cross-valid findings, indicating its effectiveness in detecting fake news. This study suggested that OPCNN-FAKE represented a state-of-the-art solution, demonstrated superior performance and promising potential for practical applications in fake news detection.

In 2020, Umer *et al.,* [4], have recommended the architecture of hybrid NN that amalgamated the capacity of LSTM and of CNN. Additionally, two distinct dimensionality reduction techniques, namely Chi-Square and Principal Component Analysis (PCA), were incorporated. Before entering the vectors of features into the classification algorithm, this work used methods of dimensionality reduction to reduce their size. A dataset with four different types of stances—disagree, discuss, unrelated, and agree—was taken from the Fake News Challenges (FNC) site in order to support the reasoning behind this technique. The chi-square and PCA were used to analyze nonlinear features, producing additional contextual information for the identification of bogus news. This study's main goal was to determine a news article's position in relation to its title. The suggested hybrid model demonstrated notable enhancements of approximately 20% and 4% in terms of F1-score and Accuracy, respectively. Experimental results underscored the superiority of PCA over state-of-the-art and Chi-Square methods, achieving an impressive 97.8% accuracy. This research contributed to advance fake news detection by leveraging the synergy of neural network architectures and dimensionality reduction techniques for more accurate and context-rich results.

In 2021, Huu *et al.,* [5], have recommended a thorough model that can accurately identify false news by taking into account both the social setting and the news content. Our method used both profound and superficial representation to explore different facets of the news material. Transformer-based methods produced deep arguments, while doc2vec and word2vec models produced shallow presentations. This representation was applied to address four distinct tasks: click bait detection, bias detection, toxicity detection, and sentiment analysis diagram. A kind of NN design called Graphical Convolution Neural Networks (GCNNs) was intended to operate with graph-structured information. By utilized

the social context data of the articles was able to strategically consider the intrinsic association between them. Experimentation on commonly-used datasets that serve as benchmarks revealed that our proposed technique was remarkably effective. The model showcased its ability to comprehensively analyze news content, and incorporated both deep and shallow representations, while leveraged the structural information through mean-field layers and GCNN. This research contributed to advancing the understanding and identification of fake news by integrated multiple facets of social context and news content data.

In 2021, Ying *et al.,* [6], have proposed a new end-to-end "Multi-modal Topic Memory Network (MTMN)" designed to acquire and amalgamate post representations shared across latent topics. This approach handled both intra- and inter-modality data within a single structure, while also included global aspects of hidden topics. MTMN's effectiveness is defined by two major breakthroughs: In real-world situations, newly arrived postings could have subject distribution that differs from those in the instruction set. This was handled by MTMN. In order to address this, we implemented a topic storage module in our technique, which defines the final image as a post feature that is shared between world and subject features of dormant topics. MTMN presented a novel blending focus module for the fusing of multi-modal data in posts, successfully integrating multi-modality data. The topic of this course utilized both the inter-modality links among text terms and image regions as well as the intra-modality linkages within each of the modes at the same time. Through mutual enhancement and complementarily, this method produced high-caliber representation. In-depth tests on two publicly available data sets proved that MTMN outperformed other cutting-edge algorithms in terms of speed. These advancements, particularly the incorporation of the topic memory module and the blended attention module, showcased MTMN's effectiveness in handling diverse topics and integrated multi-modal information for improved representation in the context of post analysis.

In 2023, Syed *et al.,* [7], have suggested to detect fake news, used a cutting-edge method that combined weakly supervised learning with machine learning to give labels to unlabeled information. The identification of fraudulent news was carried out by applying Bi-LSTM deep learning in conjunction with Bi-GRU methodologies. Techniques such as TF-IDF and Count Vectorizers were used to extract features. By integrated weakly supervised SVM techniques, the Bi-GRU deep learning and Bi-LSTM models demonstrated a remarkable

precision of 90% in finding fake news. The key innovation in this approach lied in its ability to label large volumes of unlabeled data effectively using weakly supervised learning, coupled with the efficiency of deep learning techniques in distinguishing between real and fake news when no labels was initially available. This hybrid methodology, combined weakly supervised learning and deep learning, proved to be highly effective and efficient, especially in scenarios where data lacked pre-existing labels. The achieved accuracy of 90% underscored the robustness of the approach, showcased its potential for practical application in the detection of fake news.

In 2023, Palani *et al.*, [8], have proposed the "hybrid BERT-BiLSTM-CNN (BBC-FND)" approach was structured with three key players: a layer of feature representation, classification layer and an embedding layer BERT were employed in the first layer to identify contextually-dependent characteristics between words in stories. The wideband CNN and stacking BiLSTM were both integrated into the data extractor layer. Multiple characteristics were recovered by the multichannel CNN, which captured complex local trends in word spatial interactions. Concurrently, global semantic features were retrieved from the context feature set by the stacked BiLSTM. These different traits were then combined and fed into an FFN to determine the veracity of the news. Findings showed that the BBC-FND emulate outperformed cutting-edge methods, achieving higher precision rates of 98.64%, 97.31%, 98.26% and 99.06% on four different datasets, respectively. This indicates the effectiveness of the proposed model in outperformed existing methods, highlighting its potential for robust and accurate fake news classification.

Table 2.1: Characteristics and Defects of Existing Fake News Detection Model using Deep Learning.

Author [citation]	Methodology	Features	Challenges
Abdul *et al.* [1]	CNN-RNN	• It may record the ordered as well as local properties of the input data. • It is useful for classification and regression tasks.	• It has trouble determining which hyper parameters are best for each task and data. • It required huge training dataset.
Jiang *et al.* [2]	Random forest	• The accuracy and robustness of this classifier surpass those of other classifier. • It has the highest classification performance. • It performance better with larger dataset.	• It uses multiple decision trees; hence it is slower than other classification model.
Saleh *et al.* [3]	OPCNN-FAKE	• The accuracy and robustness of this classifier surpass those of another classifier. • It extracts the high-level and low-level features.	• It requires lot of training data. • The amount of data required depends on the complexity of the task
] Umer *et al.* [4	PCA	• It provides more contextual features for detection. • It achieved better accuracy than other model.	• It leads to loss of information. • It required data normalization.
Huu *et al.* [5]	Representation learning	• It tackles four distinct tasks: sentiment analysis, click bait detection, bias detection and toxicity detection.	• It has limited representational power.
Ying *et al.* [6]	MTMN	• It improves the efficiency of fake news detection.	• It is difficult to extract complementary and comprehensive information.
Syed *et al.* [7]	Weakly supervised learning	• It accurately labels significant amounts of unlabelled data. • It is used for data annotation.	• It considers only the machine learning techniques for data annotation.
Palani *et al.* [8]	BBC-FND	• It achieves higher accuracy on four datasets • It facilitates the extraction of word attributes that depend on environment.	• Lack of ability to be spatially invariant to the input data

3. MATERIAL AND METHOD

3.1 DESIGNING OF FAKE NEWS DETECTION MODEL USING MULTI SCALE FEATURE FUSION-BASED DEEP LEARNING METHODOLOGY

3.1.1 Experimented Dataset Details

In the contemporary landscape of information dissemination through digital platforms, the surge in fake news has become a significant concern. Addressing this issue requires effective models capable of distinguishing between genuine and fabricated news content. Two datasets, namely the "Fake News Challenge" and the "Fake News Classification," have been explored for this purpose.

a. Dataset1: The “Fake News Challenge” dataset from the link of https: // github .com /FakeNewsChallenge/fnc-1/blob/master/train_stances.csv. with access: 13-11-2023. This dataset typically included various features, such as the text of the news articles, metadata, and other relevant information that could be used to train models. The goal was for participants to develop models that could generalize well to new, unseen data and accurately classify whether a news article was real or fake. The challenge aimed to address the growing concern of misinformation and fake news spreading through online platforms. By providing a standardized dataset and evaluation metrics, the Fake News Challenge encouraged the development of effective models for distinguishing between genuine and fabricated news content.

b. Dataset2: The “Fake News Classification” dataset from the link of https://www .kaggle.com/datasets/saurabhshahane/fake-news-classification with access:13-11-2023. This dataset consists of 72,134 news articles, with 35,028 labelled as real and 37,106 labelled as fake. The authors created this dataset by merging four popular news datasets: McIntire, Kaggle, BuzzFeed Political and Reuters. This merging of datasets aims to prevent overfitting of classifiers and provide a diverse set of text data for more effective machine learning training. This dataset appears to be appropriate for testing and refining machine learning algorithms for the identification of false news. Combining multiple data sets can enhance the models' capacity for variety and generalization. The gathered text data is expressed as TY_{sl}^{Rt} .

3.1.2 Brief Description of Proposed System

The contemporary era, characterized by the rapid dissemination of data across digital platforms, has witnessed a concerning surge in fake news – the intentional propagation of false or misleading data [43]. This phenomenon poses a substantial threat to public discourse, erodes trust in media, and even jeopardizes democratic processes. Conventional methods are grappling to keep pace with the ever-evolving landscape of misinformation, prompting the exploration of deep learning techniques as a promising solution in the battle against fake news.

Deep learning, a subset of artificial intelligence, showcases unparalleled prowess in deciphering intricate patterns and representations from extensive datasets [45]. This adaptability positions it as an ideal candidate for addressing the dynamic nature of fake news, which often mutates swiftly to capitalize on current events and societal tensions. This synopsis delves into the realm of applying deep learning to construct a robust and effective fake news detection system [26].

The motivation behind employing deep learning in this context stems from its inherent capacity to autonomously glean intricate features and patterns from large datasets. In contrast, traditional methods like rule-based or statistical approaches grapple with the challenge of adapting to the ever-changing strategies employed by purveyors of misinformation. Deep learning models, by contrast, have the ability to autonomously discover relevant features and evolve their understanding in response to the continuous influx of new data. However, the lack of interpretability in these models raises concerns that could undermine trust in their decision-making processes [47].

Consequently, there is a pressing need to enhance transparency and understanding in the way these models arrive at decisions [48]. The dynamic nature of fake news necessitates models that can swiftly adapt to new tactics. Designing such models, capable of continuous learning and quick adaptation to emerging patterns, is a formidable challenge that demands real-time updates and dynamic retraining. Moreover, the resource-intensive nature of collecting and annotating diverse datasets for training poses additional hurdles, especially given the dynamic nature of fake news [49].

To reduce these challenges and mitigate errors, a proposed model is put forth in this synopsis, aiming to enhance detection performance. Figure 1 provides a visual representation of the suggested detection of fake news model based on deep learning, serving as a guide for

understanding its architecture and functioning. In essence, the quest for effective detection of fake news through deep learning entails navigating a complex landscape of challenges and innovations, with the ultimate goal of preserving the integrity of information in the digital age.

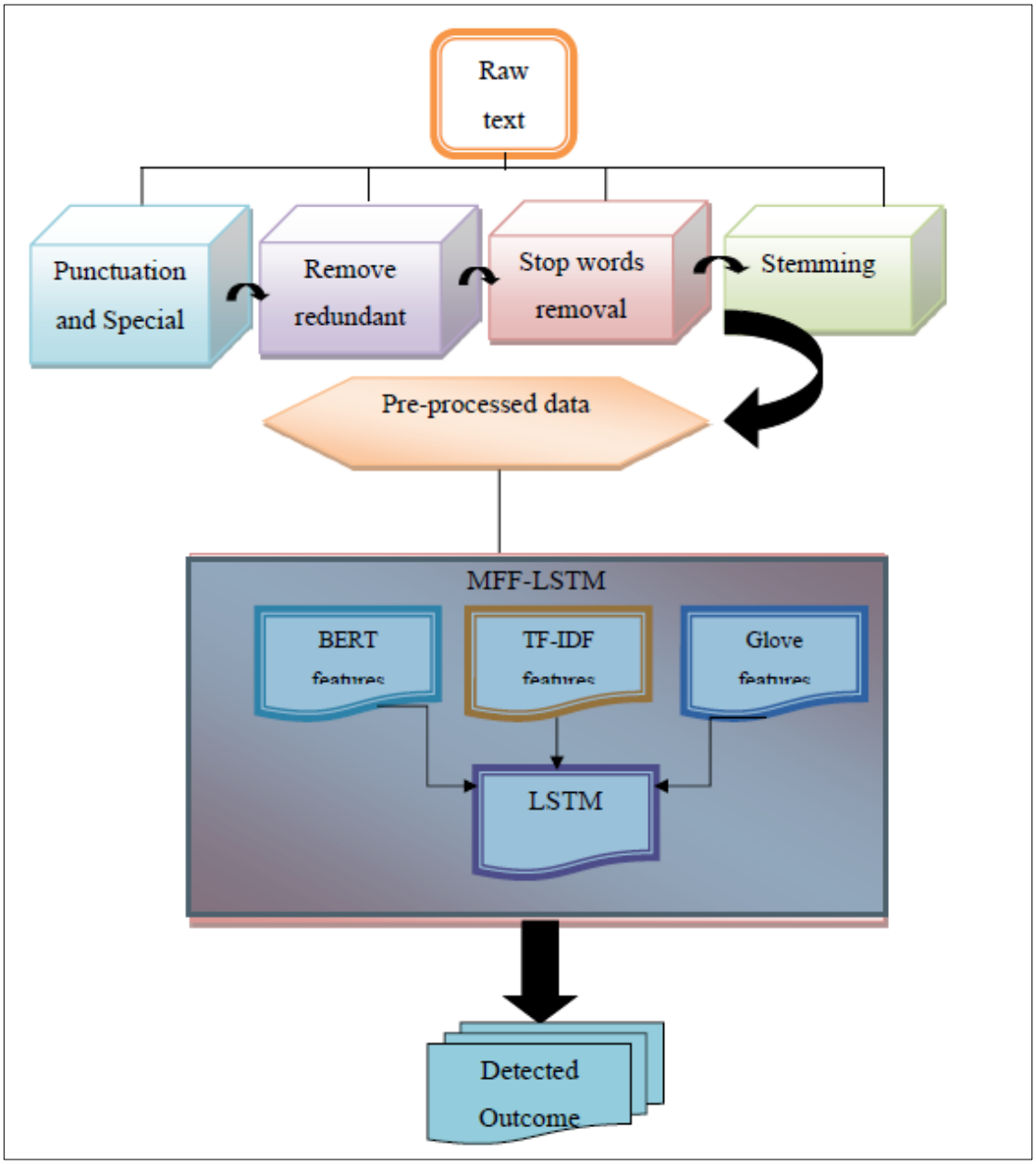

Figure 3.1 The Architecture View of The Suggested Detection Model of Fake News Based on Deep Learning.

The integrity of knowledge and public opinion are seriously threatened by the spread of disinformation in the modern digital era. It is crucial to create an adaptive approach to learning for incorrect information identification in order to overcome this difficulty. This model is designed to dynamically adjust and improve its performance based on the evolving data it encounters, aiming to stay ahead of emerging deceptive tactics. In this comprehensive approach, various techniques and technologies are integrated to create a robust system for detecting fake news with high accuracy. The initial step involves gathering input text from benchmark datasets and performing essential pre-processing tasks. This includes “removing stop words, stemming, lemmatization, and handling special characters” to prepare the text for subsequent analysis. Clean and well-processed data form the foundation for accurate and effective model training. Three distinct feature extraction techniques “BERT, TF-IDF, and GloVe Embedding” are employed. Each technique contributes a unique set of features to the model. The fusion of these features enhances the model's ability to discern patterns and relevant features crucial for fake news detection. This holistic approach reflects a comprehensive understanding of the strengths of each method. The features extracted in the previous step are combined in a multiscale manner using the MFF-LSTM model. This involves applying the extracted features to the remaining layers of the LSTM network, leveraging advanced techniques for the detection of fake news. The integration of these methods aims to enhance the model's performance in accurately identifying deceptive content within input text. To assess the effectiveness and superiority of the developed system, a crucial step involves comparing its performance with traditional approaches to fake news detection. This evaluation provides insights into the model's capability to achieve high accuracy. The ultimate goal is to create a sophisticated and adaptive learning model that significantly contributes to preventing the spread of false information and misinformation. In conclusion, this approach represents a comprehensive strategy for fake news detection, combining adaptive learning, advanced feature extraction techniques, and multiscale feature fusion. By addressing the dynamic nature of misinformation and leveraging the strengths of various methodologies, the proposed model aims to set a new standard for accuracy and efficiency in detecting fake news, contributing to the broader effort to ensure the integrity of information in the digital landscape.

3.1.3 Text Pre-processing

An essential stage in NLP is text pre-processing [31], which entails sanitizing and converting unprocessed text input into a format appropriate for evaluation, machine learning, or other NLP uses. The primary goal of text pre-processing is to convert unstructured text data into a structured and normalized format, reducing noise and improving the efficiency and effectiveness of subsequent NLP tasks. Common text pre-processing advantages and challenges are mentioned below.

a. Advantages:

a) By removing noise and irrelevant information, text pre-processing enhances the quality of the data used to train models. This can lead to improved model performance and generalization.

b) Techniques like stemming help in reducing the dimensionality of the feature space by converting words to their root form. This can be beneficial in cases where the dataset is large, and the number of unique words needs to be managed.

c) Pre-processing can make the text more interpretable by standardizing the format and removing extraneous information. This is particularly important in tasks such as topic modelling or sentiment analysis.

d) It can improve the efficiency of downstream processes by reducing the amount of data to be processed and speeding up training times.

In tasks using textual data in NLP and artificial intelligence, it is a crucial stage. It entails preparing unprocessed text data for evaluation and modelling by cleansing and formatting it. These are a few typical methods for text preparation.

a. "Punctuation and Special character removal
b. Remove redundant and inappropriate data
c. Stop words removal
d. Stemming"

3.1.3.1 Punctuation and special character removal

It [32], is a fundamental step in pre-processing text and the input of this phase is TY_{sl}^{Rt}. This technique involves the elimination of punctuation marks, symbols, and special characters from raw text data. The primary goal is to clean and simplify the text, making it more suitable for subsequent analysis, machine learning tasks, or other applications of NLP. Eventually,

the data and its representation with the punctuation and special characters removed were retrieved as TY_{pun}^{Rt}.

a. Advantages:

a) Removal of punctuation marks simplifies the tokenization process, making it easier to break text into meaningful units. This is crucial for various NLP tasks such as text classification, information retrieval and sentiment analysis.
b) Punctuation and special characters often add noise to the text, and their removal helps in creating a cleaner and more focused representation of the content. This reduction in noise can lead to more accurate analysis and modelling.

3.1.3.2 Remove redundant and inappropriate data

In text pre-processing, the removal of redundant and inappropriate data [33], is a critical step aimed at refining raw textual content for subsequent analysis and the input representation is TY_{sl}^{Rt}, machine learning, or NLP applications. Redundant and inappropriate data can include extraneous information, irrelevant details, or elements that may hinder the accuracy and effectiveness of downstream tasks. The main goal of deleting unnecessary and duplicate data is to improve the text's quality by getting rid of things that don't significantly advance the intended modelling or research. This process ensures that the resulting text data is concise, focused, and aligned with the objectives of the specific NLP task. At last obtained the redundant and inappropriate removed data and the representation is TY_{redata}^{Rt}.

a. Advantages

a) Eliminating redundant information allows the analysis or model to focus on the core content of the text, leading to more meaningful insights.
b) The removal of inappropriate data enhances the precision of subsequent tasks by ensuring that only relevant and contextually significant information is retained.

3.1.3.3 Stop words removal

These [34], are common words that frequently appear in a language but often carry little semantic meaning and may not contribute significantly to the understanding of a text. The input of this phase is TY_{sl}^{Rt}, some examples include words like "the," "and," “is," and "in." In text pre-processing, the removal of stop words is a crucial step aimed at improving the efficiency and effectiveness of subsequent NLP tasks. The primary purpose of stop words

removal is to eliminate words that are deemed non-informative, allowing the analysis or modelling process to focus on the more significant and contextually relevant terms. This stage lowers sound, boosts algorithm performance, and improves representation of features quality. Eventually, the stop word removal data was received, and the resulting representation is PR_{data}^{Rt}.

a. Advantages

a) Removing stop words reduces the number of unique words in the dataset, leading to a lower-dimensional feature space. This can be particularly beneficial in scenarios with limited computational resources.

b) By focusing on content words, the interpretability of models is often improved. Stop words removal ensures that the most meaningful and contextually relevant terms are emphasized.

3.1.3.4 Stemming

It is an NLP technique [35], that includes breaking words down to their "stem," or basic or root shape. This phase's input is the input data TY_{sl}^{Rt} , the purpose of stemming is to simplify word variations to a common base form, thereby improving text analysis and information retrieval. Stemming is commonly applied in search engines, information retrieval systems, and text mining applications. Finally obtained the pre-processed data and the represented is PR_{data}^{Rt}. This data is used for next stage of detection.

a. Advantages

a) Stemming helps in normalizing words to their root form, reducing different variations of a word to a common base. This aids in improving the accuracy and efficiency of text processing and analysis.

b) In applications such as search engines, stemming enhances information retrieval by matching different forms of a word to a common stem. This ensures that a search for one form of a word returns results containing other related forms.

3.2 EXTRACTING FEATURE TECHNIQUES OF BERT, TF-IDF AND GLOVE EMBEDDING TO DETECT THE FAKE NEWS

3.2.1 Extracting Feature Techniques of BERT, TF-IDF and Glove Embedding to Detect the Fake News

3.2.1.1 BERT-based features

The BERT [36], represents a transformative advancement in NLP. The preprocessed data PR_{data}^{Rt} is the initialization of this phase; BERT utilizes transformer encoders as a sub-structure for pre-training models on various NLP tasks. Its execution involves two key phases: pre-training for language understanding and fine-tuning for task-specific objectives. With MLM, BERT masks random words in sentences during training and reconstructs the masked words from surrounding context at the output. This enables BERT to comprehend bi-directional contexts in sentences. The NSP mechanism allows BERT to input two sentences simultaneously and determine if the second sentence follows the first. This capability facilitates the maintenance of long-distance relationships between texts.

MLM encoder, which was aided in the initial design, is a characteristic of the BERT design. This encoder is made up of an identical layered stack. Each layer is made up of two sub layers. A multi-head self-attention system makes up the first sub-layer and a straightforward position-wise entirely interconnected forward-feeding network makes up the second. Layer normalizing is implemented by applying residual hyperlinks around each of the two sub-layers. Stated differently, each sub-layer's return is expressed as $LN(y + Sublayer(y))$, where $Sublayer(y)$ denotes the function implemented by the respective sub-layer.

The set of inquiries' values and keys is given a term l, s, m. By calculating the dots in the product of the questions and answers and scaling it by the square roots of the questions' dimension, the scaling dot-product attentiveness method calculates one's attention scores. The final product is then produced by weighing the variables using attention rate scores. The enlarged dot-product close attention is represented scientifically by Eq. (1).

$$Atte(l, s, m) = soft\max\left(\frac{Ls^{t}}{\sqrt{f_k}}\right)K \tag{3.1}$$

With this approach, attention scores are guaranteed to be less susceptible to variations in the search query and keyword vector scales and to be more consistent. The weights that are used for each of the values in the finished product are determined by the SoftMax function, which generates a probabilistic distribution over the keys in question.

An essential phase in the self-attention process is the scaling dot-product focus, which enables models to successfully capture connections among various words in an order. The Multi-head attention $Mulhe$ is expressed mathematically as Eq. (2).

$$Mulhe(l,s,m) = con(h_1, h_2 ..... h_n)E_s \tag{3.2}$$

The initial token in the sequence is denoted as "[CLS]". In cases where there are two sentences, they are distinguished by the SEP token within the sequence. Additionally, an embedding is appended to each token to specify whether it pertains to the first or second sentence.

Fine-Tuning: It is streamlined because the self-attention mechanism in the transformer architecture is inherently flexible, allowing BERT to effectively model various downstream tasks. In the fine-tuning phase for each specific task, seamlessly integrate task-specific outputs and inputs into BERT and subsequently fine-tune all the model parameters. Finally the BERT based feature is acquired and it is represented as $BERT_{data}^{Rt}$. Figure 2 show the diagrammatic view of BERT based Feature.

a. Advantages

a) BERT excels in capturing bi-directional contexts within sentences. Unlike traditional models that process text sequentially, BERT considers both preceding and following words for each word in a sentence. This bi-directional context contributes to a more comprehensive understanding of language and the relationships between words.

b) It is possible to refine BERT's trained models for a variety of NLP tasks. Its flexibility to various comprehension aims is demonstrated by its versatility, which makes it a strong tool for a wide range of applications, including recognition of named entities and sentiment estimation.

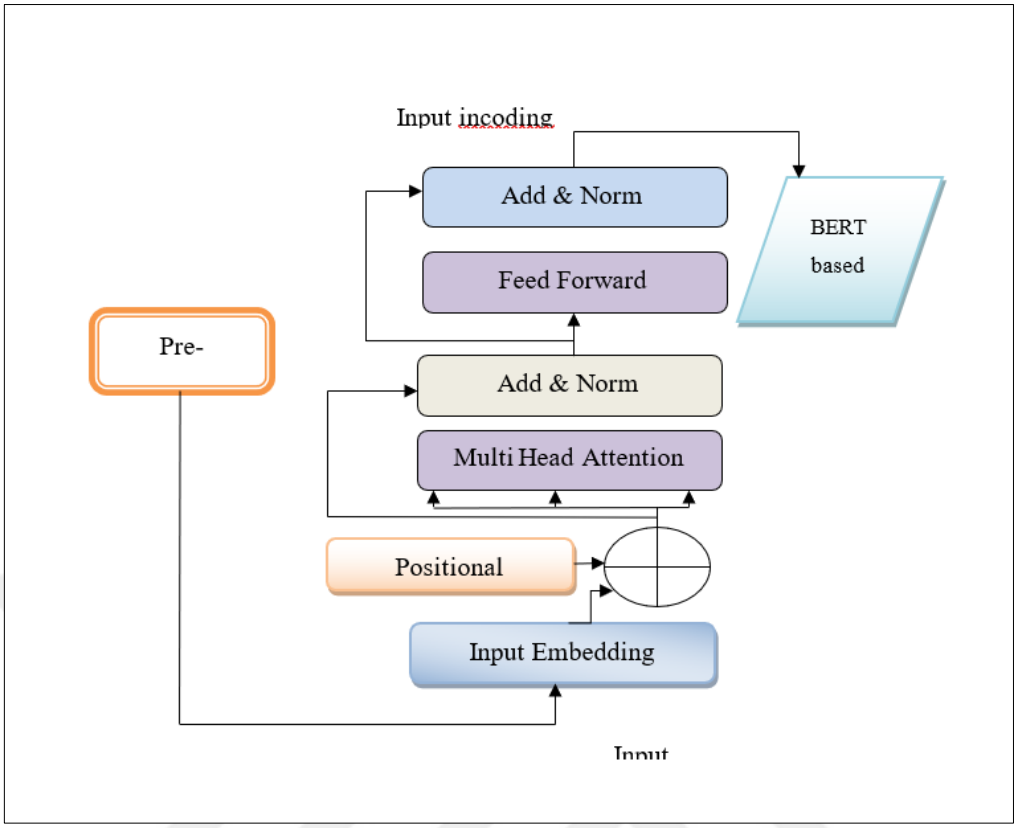

Figure 3.2: Diagrammatic view of BERT Based Feature.

3.2.1.2 TF-IDF-based features

TF-IDF [37], is a statistical measure used in NLP and information retrieval to evaluate the importance of a word in a document relative to a collection of documents, often referred to as a corpus.

The preprocessed data PR_{data}^{Rt} is the input to this phase; TF-IDF is employed to generate features that represent the significance of words in text data. These features are widely used in various NLP applications, including text classification, information retrieval, and document similarity analysis.

TF-IDF, an weighting scheme of unsupervised term extensively used in text mining and data retrieval, quantifies the significance of a variable t in a given text document. The term's relative frequency, denoted as TF, reflects how often the word appears in the document. Simultaneously, the inverse document frequency, IDF, considers the scale of the term across the entire collection of documents, which is mathematically modeled in Eq. (3).

$$K_{s,d} = ip \times \log\left(\frac{M}{jp}\right) \tag{3.3}$$

The weight assigned to a variable i in a document d is expressed by M. In the context of a corpus, where $K_{s,d}$ represents the total amount of documents, ip signifies the term frequency, indicating how many times a particular term jp denotes the document frequency, representing the number of documents containing the term.

The primary purpose of TF-IDF is to quantify the relevance of a term within a document in the context of a larger corpus. It helps address the challenge of identifying important words that contribute to the overall meaning of a document while discounting common terms that may not carry significant semantic weight. Finally got the TF-IDF based feature and the mathematical representation are $TF - IDF_{data}^{Rt}$. Figure 3 shows the diagrammatic view of TF-IDF based Feature.

a. Advantages of TF-IDF

a) TF-IDF is highly effective in quantifying the significance of terms within documents. It achieves this by considering both the local importance, which is the term frequency within a specific document, and the global importance, which is the inverse document frequency across the entire corpus. This dual consideration provides a robust measure of a term's importance.

b) TF-IDF aids in discriminating between terms that are common across many documents and those that are specific to a particular document. Common terms receive lower TF-IDF scores, while unique or rare terms receive higher scores. This helps highlight the distinctive terms that contribute to the uniqueness of a document.

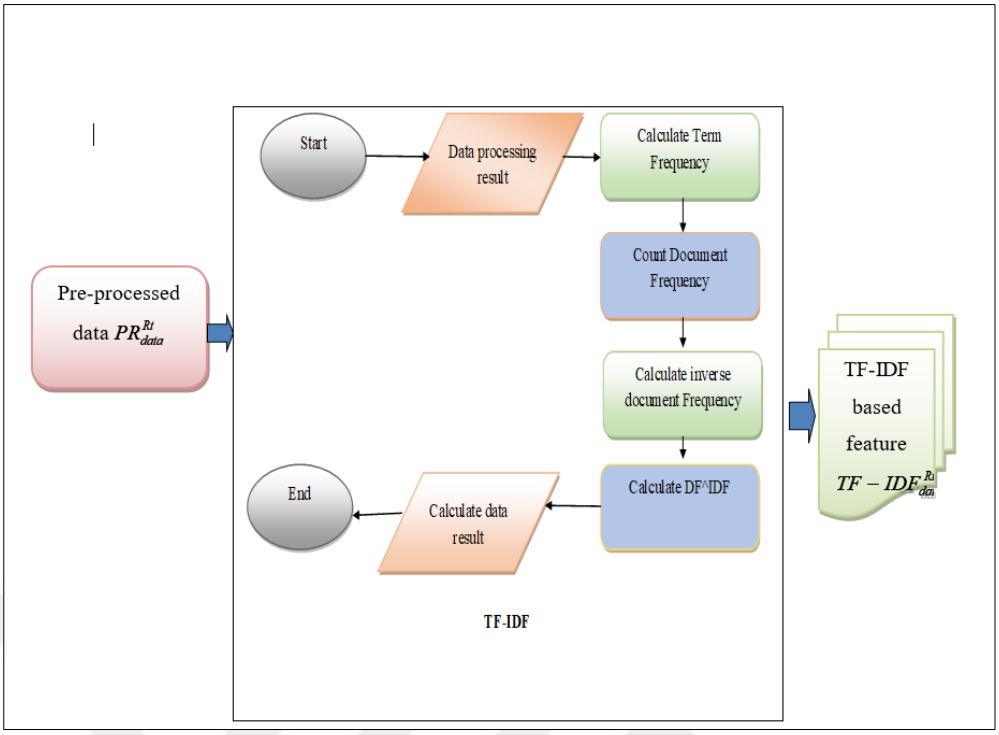

Figure 3.3: Diagrammatic View of TF-IDF Based Feature.

3.2.1.3 GloVe-based features

GloVe [38], is an unsupervised learning algorithm designed to generate word embeddings based on global statistical information. The pre-processed data PR_{data}^{Rt} is the input to this phase, the embeddings produced by GloVe capture semantic relationships between words, and these embeddings can be used to create feature vectors for words or documents. GloVe-based features are particularly useful in NLP tasks that require understanding of word semantics and relationships.

GloVe, a word2vec-based model for efficient word representation learning from text documents, was employed in this study.

To enhance the efficiency of the proposed method, TF-IDF with GloVe models was incorporated. Let $U(u_1, u_2 ....u_k)$ represent each input k-word, with each word being transformed into a d-dimensional word vector. Consequently, Rd denotes the dimension space of every variable. Thus, each input text is expressed as $sl \times e$ dimension space, and the input text matrix is generated as $U(u_1, u_2 ....u_k) \in sl \times e$. Finally, the vector of feature Re for the document K is formed by concatenating it with the word embeddings, denoted as $\oplus$, resulting in the expression is in Eq. (4).

$$\mathrm{Re} = k_1 \oplus k_2 ..... \oplus k_{n-1} \tag{3.4}$$

To enhance the quality of text representation, combine pre-trained GloVe word embeddings with TF-IDF weighting, that is mathematically explained in Eq. (5).

$$Y_i = K_{i,d} \times T_s \tag{3.5}$$

As mentioned earlier, the matrix T_s represents word vectors obtained from Glove, and $K_{i,d}$ corresponds to the TF-IDF document weight d and variable i. This approach effectively addresses the dimensional challenges associated with sparse matrices of high-dimensional.

The use of Gaussian Dropout and Gaussian Noise serves as a regularization technique, enhancing the robustness of the model and reducing the risk of over fitting. Specifically, the application of Gaussian noise directly to the word embeddings introduces a random data augmentation aspect during training.

This stochastic process contributes to the model's adaptability by introducing variability, thereby preventing it from relying too heavily on specific patterns in the training data. This regularization strategy is instrumental in improving the model's generalization performance, ensuring its efficacy across diverse datasets. Finally obtained the GloVe based feature and the representation is $Glove_{data}^{Rt}$ .Figure 4 shows the view of GloVe based Feature.

a. Advantages of GloVe

a) GloVe-based features excel in capturing semantic relationships between words. This enables a more nuanced understanding of word meanings by representing them in a continuous vector space. The model's ability to discern semantic nuances contributes to a richer representation of language.
b) Leveraging global co-occurrence statistics, GloVe considers the broader context of word relationships across the entire corpus. This holistic approach provides a comprehensive view of word semantics, taking into account the relationships that words share on a global scale within the given dataset.
c) GloVe embeddings typically have a lower-dimensional representation compared to one-hot encodings or some other embedding methods. This reduction in dimensionality can lead to more computationally efficient models and improved generalization performance.

d) GloVe embeddings have become widely adopted in the NLP community and are used as a foundational component in various applications, including sentiment analysis, machine translation, and information retrieval.

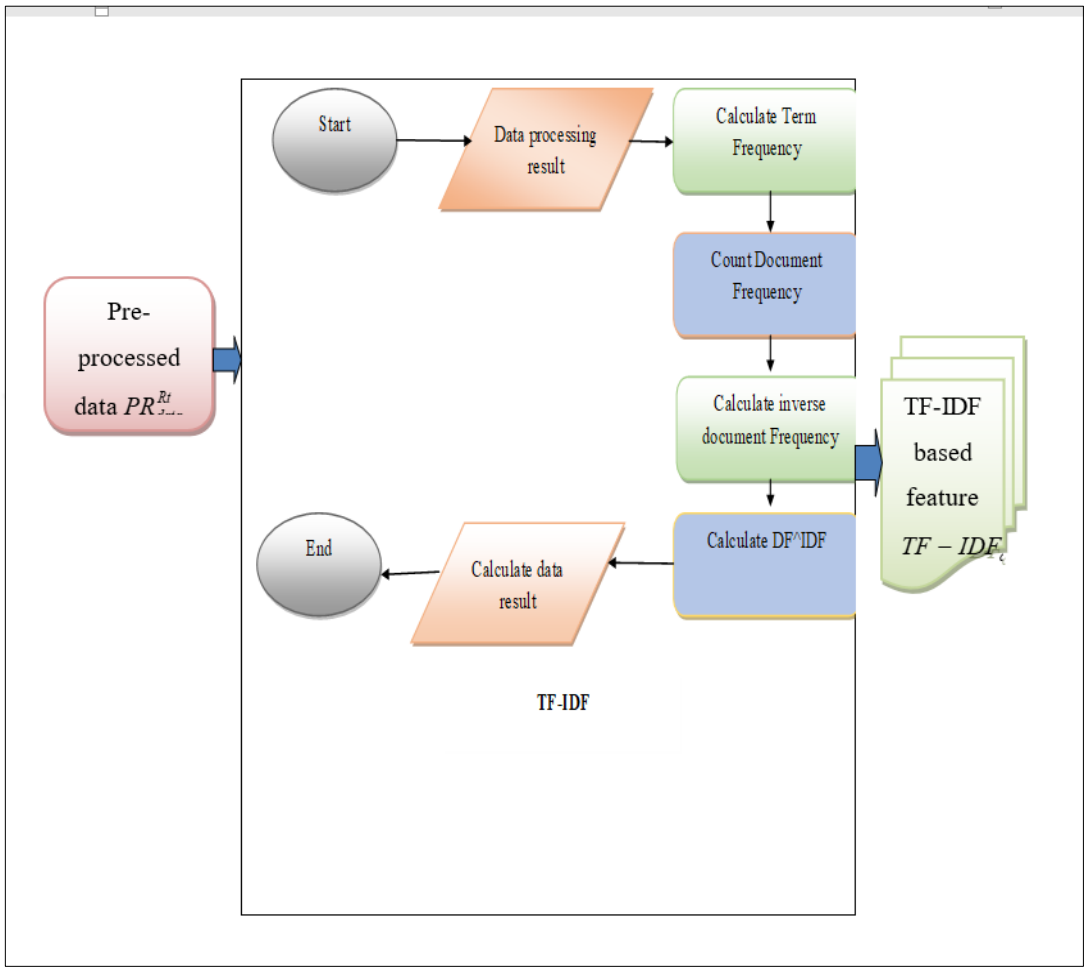

Figure 3.4: Diagrammatic View of BERT Based Feature.

3.2.1.4 GloVe-based features

GloVe [38], is an unsupervised learning algorithm designed to generate word embeddings based on global statistical information. The pre-processed data PR_{data}^{Rt} is the input to this phase, the embeddings produced by GloVe capture semantic relationships between words, and these embeddings can be used to create feature vectors for words or documents. GloVe-based features are particularly useful in NLP tasks that require understanding of word semantics and relationships.

GloVe, a word2vec-based model for efficient word representation learning from text documents, was employed in this study.

To enhance the efficiency of the proposed method, TF-IDF with GloVe models was incorporated. Let $U(u_1, u_2....u_k)$ represent each input k-word, with each word being transformed into a d-dimensional word vector. Consequently, Rd denotes the dimension space of every variable. Thus, each input text is expressed as $sl \times e$ dimension space, and the input text matrix is generated as $U(u_1, u_2....u_k) \in sl \times e$. Finally, the vector of feature Re for the document K is formed by concatenating it with the word embeddings, denoted as $\oplus$, resulting in the expression is in Eq. (4).

$$\mathrm{Re} = k_1 \oplus k_2 ..... \oplus k_{n-1} \tag{3.6}$$

To enhance the quality of text representation, combine pre-trained GloVe word embeddings with TF-IDF weighting, that is mathematically explained in Eq. (5).

$$Y_i = K_{i,d} \times T_s \tag{3.7}$$

As mentioned earlier, the matrix T_s represents word vectors obtained from Glove, and $K_{i,d}$ corresponds to the TF-IDF document weight d and variable i. This approach effectively addresses the dimensional challenges associated with sparse matrices of high-dimensional.
The use of Gaussian Dropout and Gaussian Noise serves as a regularization technique, enhancing the robustness of the model and reducing the risk of over fitting. Specifically, the application of Gaussian noise directly to the word embeddings introduces a random data augmentation aspect during training.
This stochastic process contributes to the model's adaptability by introducing variability, thereby preventing it from relying too heavily on specific patterns in the training data. This regularization strategy is instrumental in improving the model's generalization performance, ensuring its efficacy across diverse datasets. Finally obtained the GloVe based feature and the representation is $Glove_{data}^{Rt}$ .Figure 4 shows the view of GloVe based Feature.

a. Advantages of GloVe

a) GloVe-based features excel in capturing semantic relationships between words. This enables a more nuanced understanding of word meanings by representing them in a continuous vector space. The model's ability to discern semantic nuances contributes to a richer representation of language.

b) Leveraging global co-occurrence statistics, GloVe considers the broader context of word relationships across the entire corpus. This holistic approach provides a comprehensive view of word semantics, taking into account the relationships that words share on a global scale within the given dataset.
c) GloVe embeddings typically have a lower-dimensional representation compared to one-hot encodings or some other embedding methods. This reduction in dimensionality can lead to more computationally efficient models and improved generalization performance.
d) GloVe embeddings have become widely adopted in the NLP community and are used as a foundational component in various applications, including sentiment analysis, machine translation, and information retrieval.

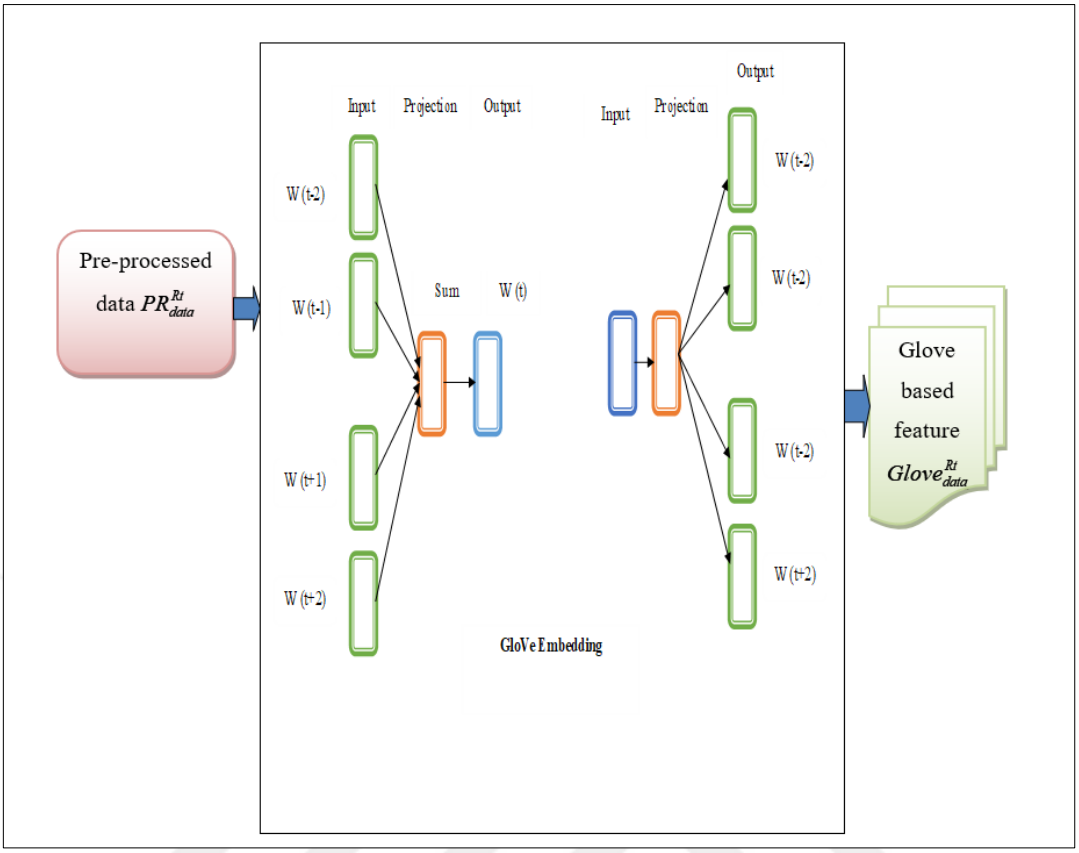

Figure 3.5: Diagrammatic View of GloVe Based Feature.

3.3 FAKE NEWS DETECTION BY APPLYING THE MULTI-SCALE FEATURE FUSION-AIDED LONG SHORT-TERM MEMORY

3.3.1 Multi-scale Concept and its Benefits

The concept of multi-scale [39], in deep learning involves incorporating information from various resolutions or scales within a neural network. This strategy entails handling input data or features at different levels of detail, enabling the model to comprehend information that spans from intricate details to broader patterns [50][51]. Multi-scale architectures are commonly found in diverse deep learning models and have demonstrated advantages across a spectrum of tasks. The benefits of multiscale concept are described below.

In our proposed model, the obtained three features $BERT_{data}^{Rt}$, $TF-IDF_{data}^{Rt}$ and $Glove_{data}^{Rt}$ are concatenated and finally obtained the extracted feature, which is represented as EF_{data}^{Rt}.

a. Multi-scale processing makes the deep learning models as more robust to variations in the scale or size of objects within the input data. This is particularly beneficial in computer vision tasks where objects may appear at different scales.
b. By incorporating information from different scales, models gain a more comprehensive understanding of the context in which features exist. This is crucial for tasks where both local and global context contribute to the overall understanding of the input.
c. Multi-scale architectures enable the extraction of features at various levels of abstraction. This results in more informative feature representations that capture both intricate details and high-level patterns [52].
d. Deep learning models with multi-scale capabilities often perform well on diverse inputs, accommodating a range of structures and patterns. This adaptability is valuable in scenarios where input data exhibit variability in terms of size, shape, or complexity.
e. Multi-scale architectures can be designed to use computational resources efficiently. Some layers or pathways may operate at lower resolutions, reducing the computational burden while preserving valuable information.

In summary, the multi-scale concept in deep learning is a versatile approach that enhances model performance by considering information at multiple resolutions. Its benefits include robustness to scale variations, improved contextual understanding, and more effective feature representation, making it applicable to a wide range of tasks across various domains.

3.3.2 Basic LSTM

LSTM was developed as an improved version of the standard RNN with the purpose of controlling and managing the information flow within the network. Its primary objective is to address issues such as the vanishing and exploding gradient problems commonly encountered in standard RNNs, which can hinder effective learning over long sequences.

LSTM [29][53], achieves this by enhancing the memory capacity of RNNs. Throughout the training process; it effectively handles error propagation for back-propagation through multiple layers, enabling learning to occur over extended sequences [54]. The key innovation of LSTM lies in its ability to capture long-distance dependencies within sequential data, thereby retaining contextual semantic data over extended contexts using specialized memory cells.

Each LSTM unit comprises an output gate, input gate and forget gate which collectively coordinate the decision-making process regarding the information to be discarded, passed or retained on to the next step.

These gates play a crucial role in determining when to allow write, delete operations or read effectively controlling the flow of information within the LSTM unit. The computation of the forget, output and input gates along with the input cell state, is governed by Eq. (6) to Eq. (9).

These equations define the mechanisms through which LSTM units decide what information to remember, forget, and output, contributing to the model's ability to capture and learn from long-range dependencies in sequential data.

$$T_t = \sigma(e_T p_t + e_s p_{t-1} + R_T) \tag{3.8}$$

$$F_t = \sigma(e_F p_t + e_s p_{t-1} + R_F) \tag{3.9}$$

$$O_t = \sigma(e_O p_t + e_s p_{t-1} + R_O) \tag{3.10}$$

$$K_t = tahk(Y_t) \otimes Q_t \tag{3.11}$$

In the LSTM architecture, the notation $\otimes$ represents element-wise multiplication, and R_T, R_F and R_O represent bias vectors. The activation functions used are $tahk$ for hyperbolic tangent and σ for the function of sigmoid, which serves as the gate activation function. The weighing factors e_T, e_F and e_O is employed to map the three gates and input cell state to the input hidden layers.

The forget gate, output gate and input gate are the three gates that define the LSTM design. When determining how much data to keep, delete, or move on to the next stage, these gates are essential. The input gate regulates the flow of fresh data, the forget gate handles the deletion or keeping of outdated data, and the gate that delivers the result chooses which data should be sent to the following stage.

Through these gates, LSTM units effectively regulate the flow of information, allowing for the retention of important contextual information and addressing challenges posed by long-term dependencies in sequential data. The LSTM equations, incorporating the mentioned elements and functions, govern the computation of the input, forget, and output gates, as well as the input cell state. These equations contribute to the overall functionality of LSTM,

enabling it to learn and retain information over extended sequences in sequential data. Figure 5 shows the architecture view of LSTM.

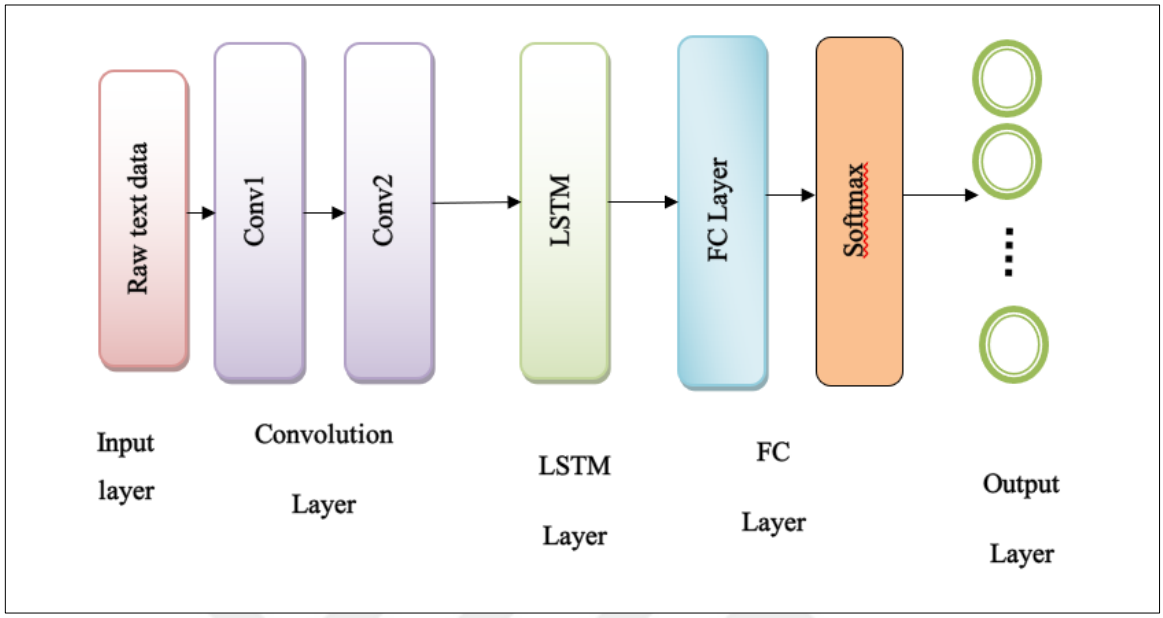

Figure 3.6: The Architecture View of LSTM.

MFF-LSTM represents a specialized approach for fake news detection, involving the integration of features at multiple scales through LSTM networks. The MFF-LSTM methodology combines features derived from various scales using LSTM networks specifically designed for the task of detecting fake news. The goal is to increase the model's ability to recognize complex connections and trends in textual data, which will ultimately increase the precision of detecting false news. The basic architecture of our proposed method is LSTM, the obtained three features $BERT_{data}^{Rt}$, $TF-IDF_{data}^{Rt}$ and $Glove_{data}^{Rt}$ is fed as input to the LSTM model, then concatenated the features and finally obtained the extracted feature. The extracted feature is fed as input to the next layer of LSTM and finally obtained the classified outcome. Figure 6 shows the proposed view of MFF-LSTM for fake news detection.

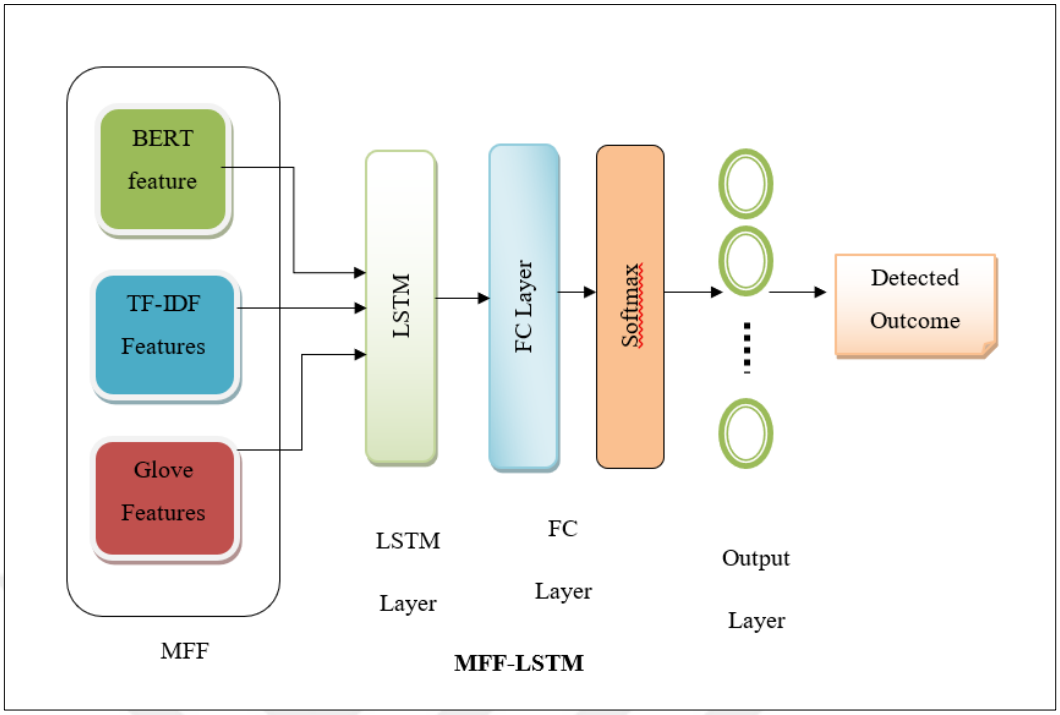

Figure 3.7: A Proposed View of MFF-LSTM For Fake News Detection.

4. RESULTS AND DISCUSSION

The spread of false news presents a serious threat to the integrity of information systems in a time when information is disseminated quickly through the internet. A deep learning-powered sophisticated false news identification system was painstakingly created in order to solve this problem. Using Python, a flexible and popular programming dialect for artificial intelligence applications, this method aims to improve the precision and effectiveness of identifying bogus news. The principal aim of this undertaking is to demonstrate the efficacy of the developed detection technique. To achieve this, the results obtained from the deep learning-based MFF-LSTM scheme are systematically compared with those derived from other well-established classification techniques. By juxtaposing these outcomes, a comprehensive assessment of the system's performance is undertaken.

The classification techniques chosen for benchmarking include Residual Attention Network (RAN) [26], Residual Network (ResNet) [27], ShufflenetV3 [28], and LSTM [29]. Each of these benchmarks represents a distinct approach to information classification, and their inclusion in the evaluation process serves to provide a holistic perspective on the capabilities of the MFF-LSTM-based scheme. The findings holded the potential to not only advances the field of fake news detection but also to inform the development of more resilient and adaptive systems for safeguarding the authenticity of information in the digital age.

4.1 PERFORMANCE METRICS

Using a variety of quantitative metrics described below, the performance of the proposed fake news detection system is assessed in order to determine the quality of the model that was created.

(a) MCC: The MCC takes into account "true positives uu_p, true negatives uu_n, false positives yy_p, and false negatives yy_n "from a confusion matrix to calculate a coefficient that ranges from -1 to 1. MKK is estimated the Eq. (10).

$$MKK = \frac{uu_p \times uu_n - yy_p \times yy_n}{\sqrt{(uu_p + yy_p)(uu_p + yy_n)(uu_n + yy_p)(uu_n + yy_n)}} \tag{4.1}$$

(b) Specificity: It focuses on the model's ability to avoid misclassifying negatives. $Spec$ is estimated in Eq. (11).

$$Spec = \frac{uu_N}{uu_N + yy_p} \tag{4.2}$$

(c) NPV: NPV represents the proportion of actual negatives among the instances predicted as negative by the model. It focuses on the model's ability to correctly predict negative outcomes. N_{pv} is estimated in Eq. (12).

$$N_{pv} = \frac{uu_n}{uu_n + yy_n} \tag{4.3}$$

(d) F1-score: It is particularly useful when there is an imbalance between the classes. $Fcore$ is estimated in Eq. (13).

$$Fcore = 2 \times \frac{2uu_P}{2uu_p + yy_p + yy_n} \tag{4.4}$$

(e) FNR: FPR, measures the proportion of actual negatives incorrectly predicted as positives by the model. $Ffnr$ is estimated in Eq. (14).

$$Ffnr = \frac{yy_n}{yy_n + uu_p} \tag{4.5}$$

(g) Sensitivity: Sensitivity measures the proportion of actual positives correctly identified by the model out of the total actual positives. Sen is estimated in Eq. (15).

$$Sen = \frac{uu_p}{uu_p + yy_n} \tag{4.6}$$

(h) FPR: It quantifies the proportion of actual negative instances that are incorrectly predicted as positive by the model. $Ffpr$ is estimated in Eq. (16).

$$Ffpr = \frac{yy_p}{yy_p + uu_n} \tag{4.7}$$

(i) FDR: FDR measures the proportion of predicted positives that are false positives. It is estimated in Eq. (17).

$$Ffdr = \frac{yy_p}{yy_p + uu_p} \tag{4.8}$$

4.2 ROC ANALYSIS FOR DATASET 1 AND DATASET 2

The fake news detection approach was assessed through the estimation of ROC against various traditional classifiers, as illustrated in Figure 7. The validation of the implemented fake news detection model was based on FPR values, revealing superior ROC rates compared to several classifiers. Specifically, it outperformed RAN by 18%, ResNet by 10.6% and ShufflenetV3 by 14.3% and LSTM by 12.6% when the FPR value is set at 0.8.

These results provide evidence of the effectiveness and superiority of the recommended fake news detection model.

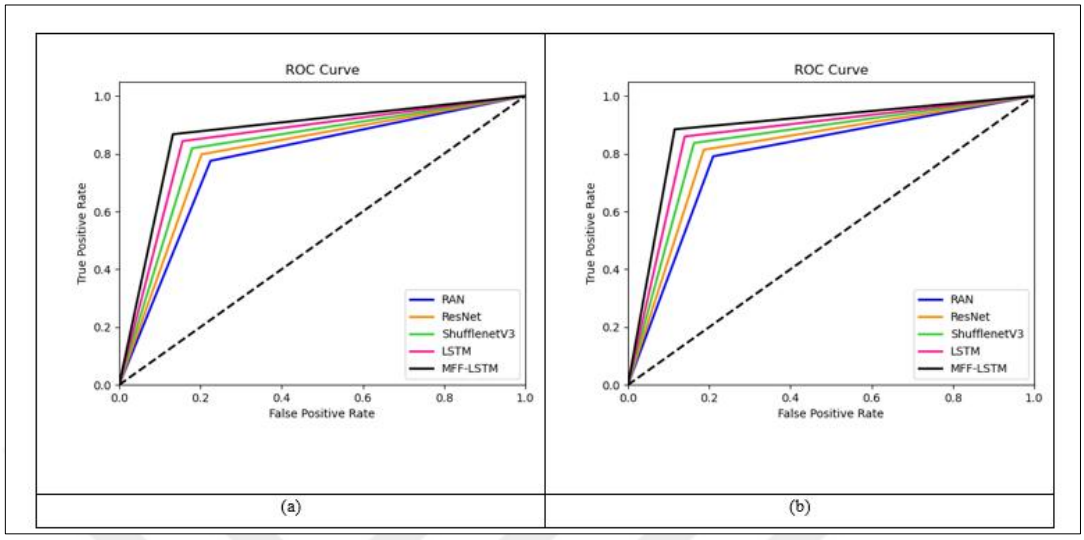

Figure 4.1: The ROC Estimation of the Developed Fake News Detection Model Over Numerous Traditional Classifiers Regarding "(a)Dataset 1 and (b) Dataset 2".

4.3 COMPARATIVE ANALYSIS OF BATCH SIZE OVER TRADITIONAL CLASSIFIER IN DATASET 1 AND DATASET 2

The efficiency of the suggested fake news detection model was evaluated across various conventional classifiers in dataset 1 and dataset 2, as depicted in Figure 8 and Figure 9. The assessment considered different batch size values for the presented work. In Figure 8 (a), when the batch size is set to 3, the suggested fake news detection model exhibited enhanced accuracy compared to RAN by 17%, ResNet by 16.6%, ShufflenetV3 by 21.3%, and LSTM by 42.6%, respectively. This underscored the superior quality of the designed detection of fake news model compared to other conventional classifier models.

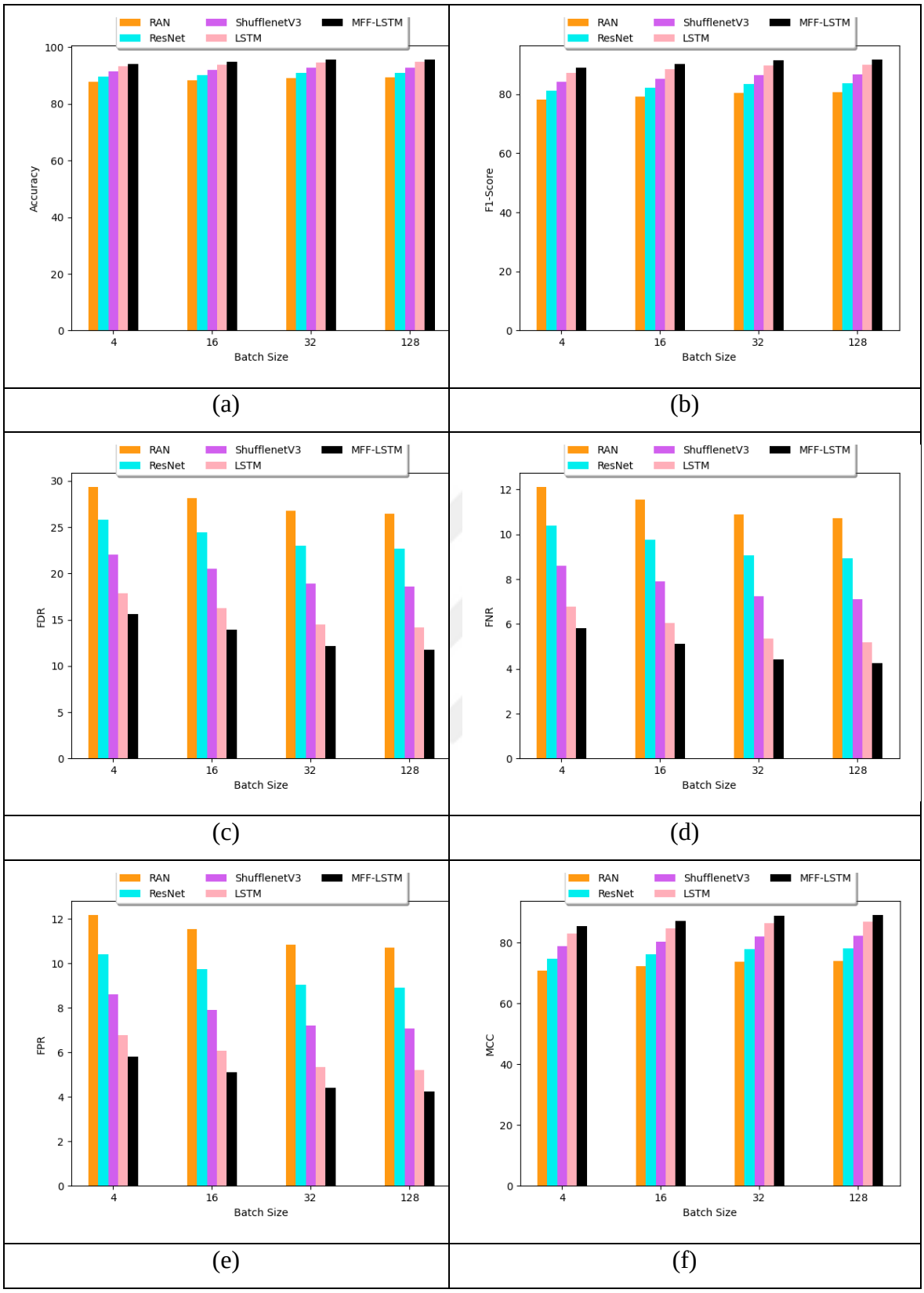

Figure 4.2: The Batch Size Evaluation of the Suggested Fake News Detection Model Over Diverse Conventional Classifier for Dataset 1 Concerning " (a) Accuracy, (b) F1-score, (c) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NPV, (h) Precision, (i) Sensitivity".

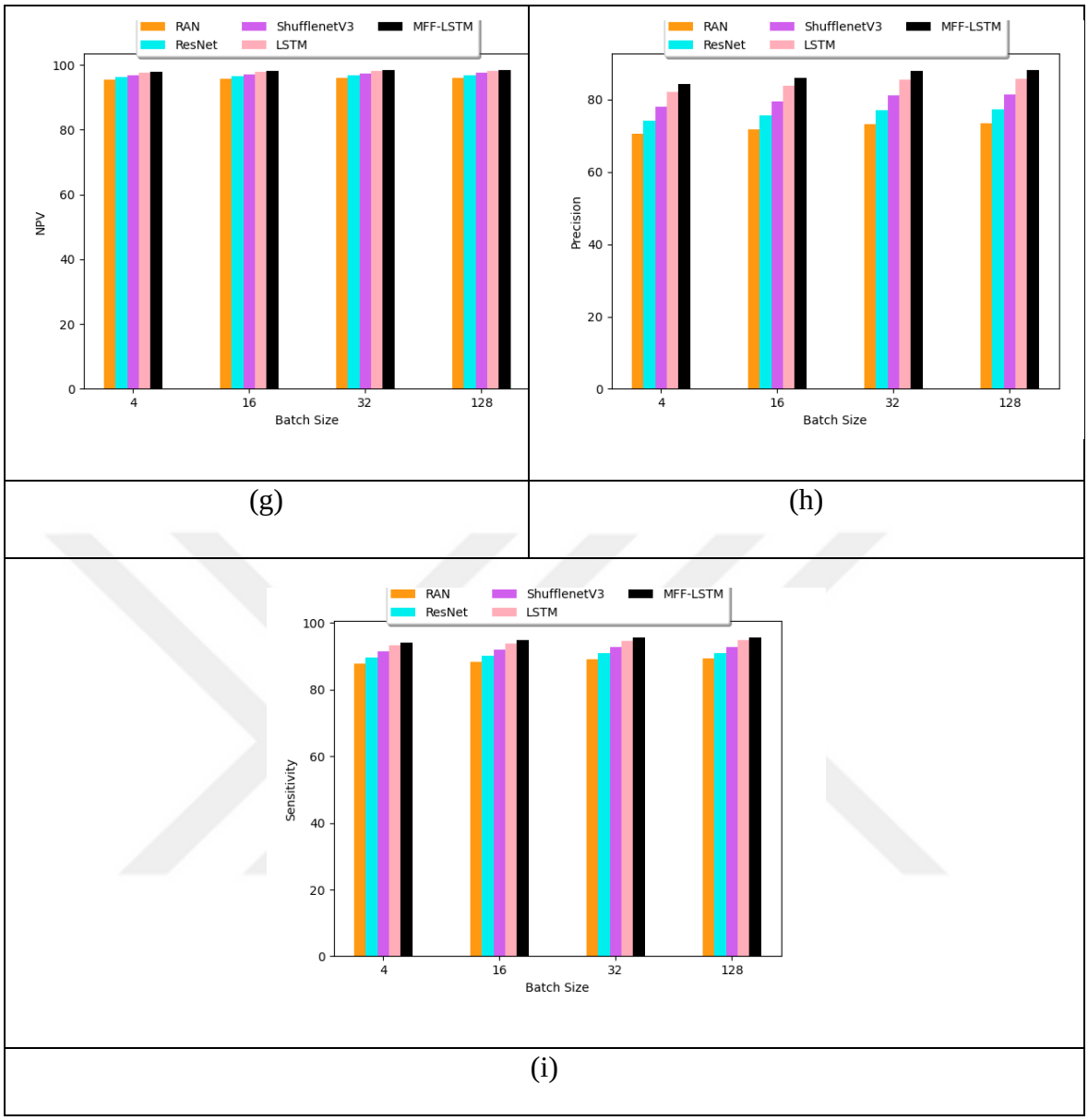

Figure 4.2: The Batch Size Evaluation of the Suggested Fake News Detection Model Over Diverse Conventional Classifier for Dataset 1 Concerning " (a) Accuracy, (b) F1-score, (c) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NPV, (h) Precision, (i) Sensitivity"."Figure Continued".

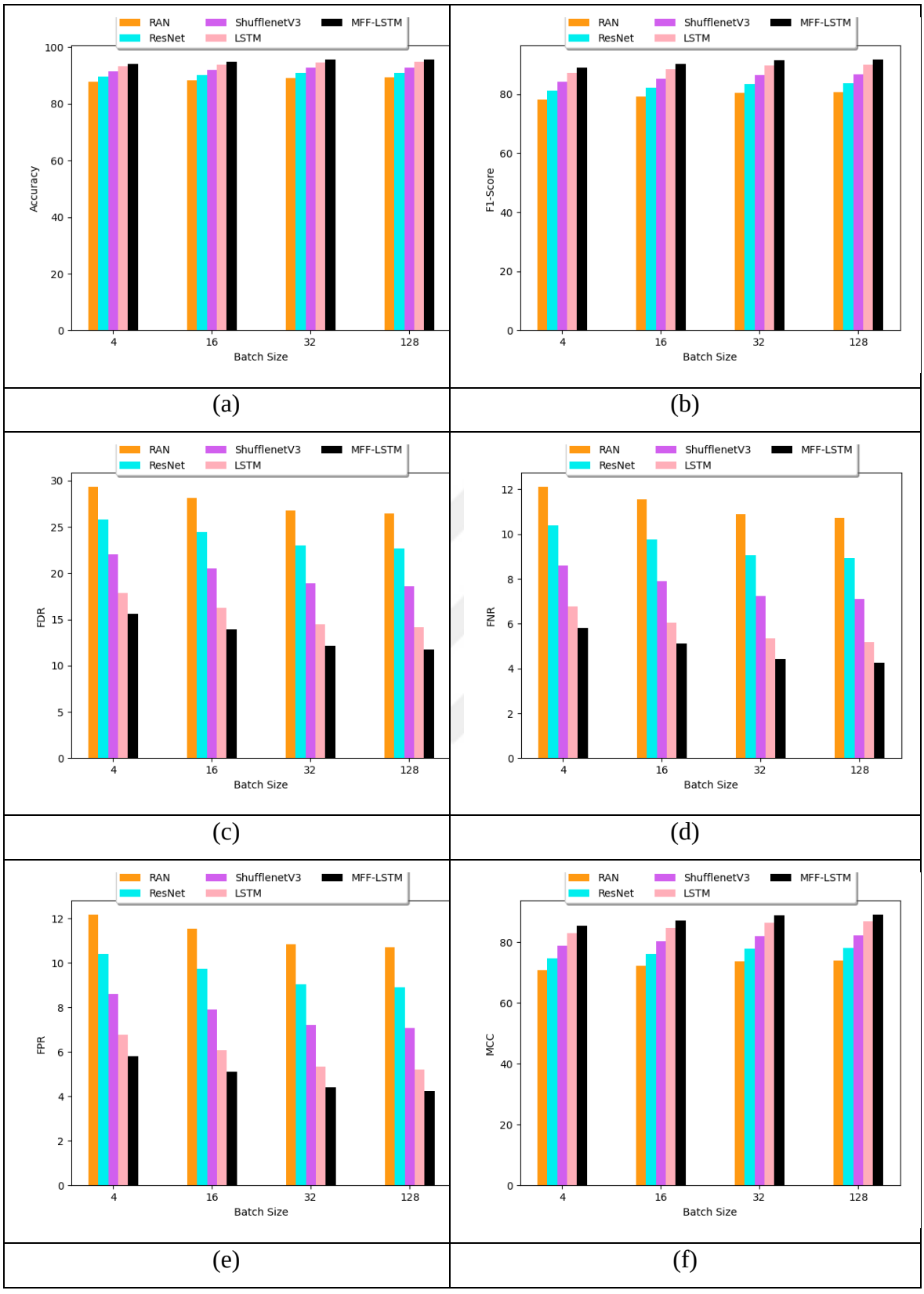

Figure 4.3: The Batch Size Evaluation of the Suggested Fake News Detection Model Over Diverse Conventional Classifier for Dataset 1 Concerning " (a) Accuracy, (b) F1-score, (c) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NPV, (h) Precision, (i) Sensitivity".

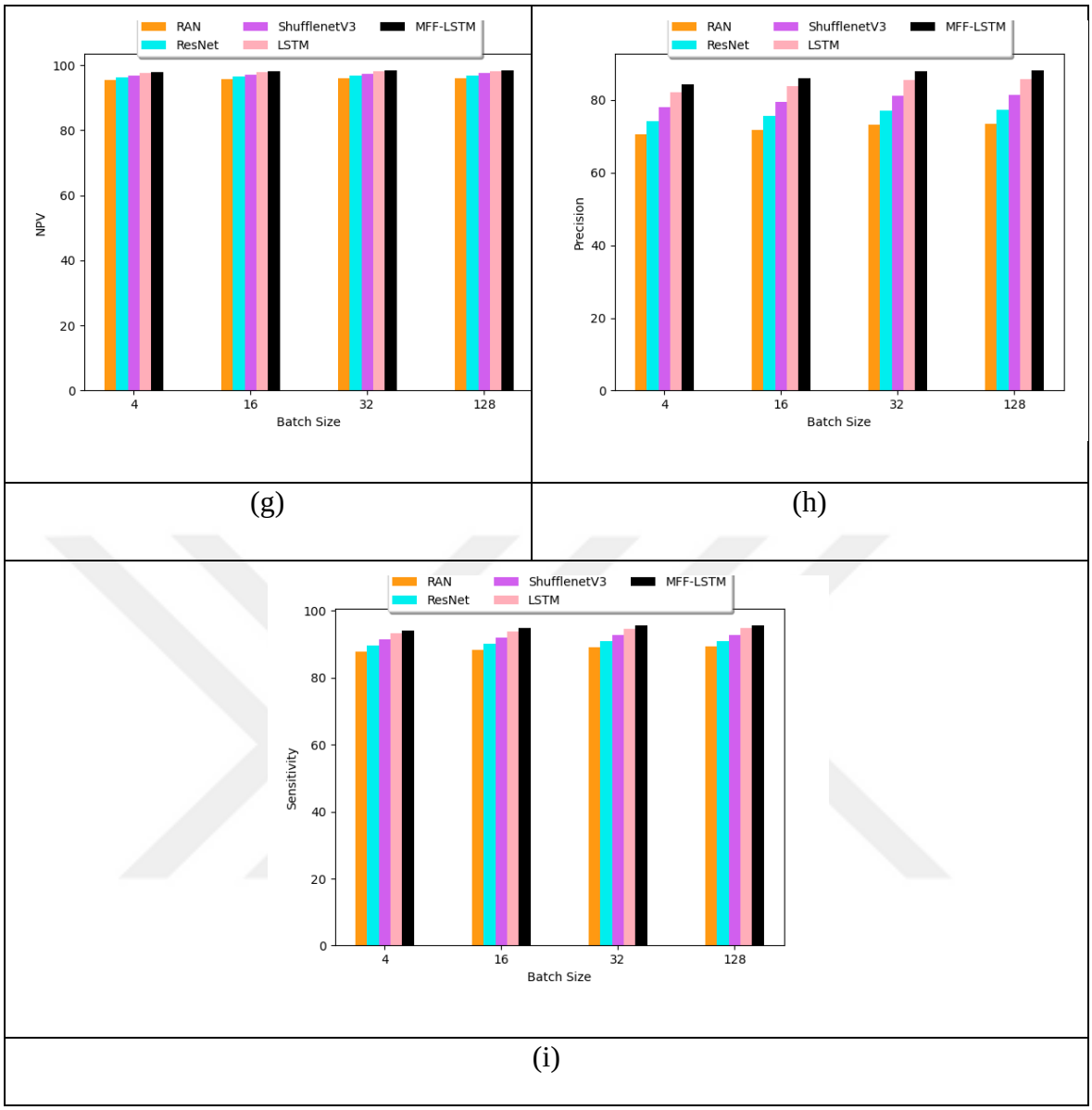

Figure 4.3: The Batch Size Evaluation of the Suggested Fake News Detection Model over Diverse Traditional Classifier for Cataset 2 Concerning " (a) Accuracy, (b) F1-score, (c) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NPV, (h) Precision, (i) Sensitivity."Figure Continoued"

4.4 PERFORMANCE ANALYSIS ON FEATURE EXTRACTION FOR DATASET 1 AND DATASET 2

The feature extraction of the developed fake news detection model was evaluated across various traditional classifiers in dataset 1 and dataset 2, as depicted in Figure 10 and Figure 11. The assessment considered different feature extracted techniques for the presented work. In Figure 10 (a), when the technique is set to glove feature, the suggested fake news detection model exhibited enhanced accuracy compared to RAN by 45%, ResNet by 34%, and Shufflenet V3 by 67%, and LSTM by 57%, respectively. This underscored the superior quality of the designed fake news detection model's feature compared to other traditional classifier models.

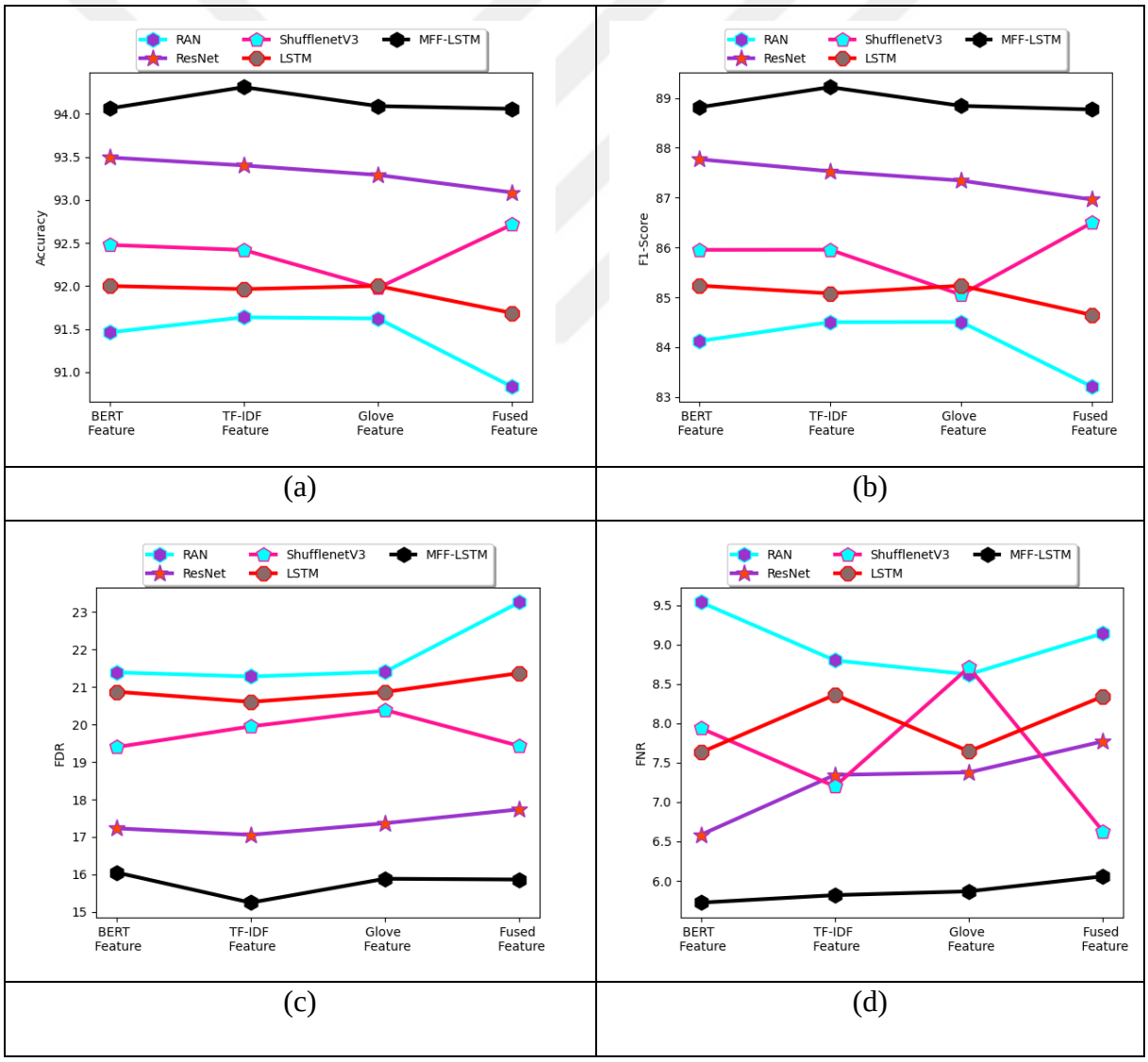

Figure 4.4: Figure 10:Comparative Evaluation of Recommended Fake News Detection Model's Feature Extraction over Diverse Conventional Classifier for Dataset 1 Concerning " (a) Accuracy, (b) F1-score, (c) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NPV, (h) Precision, (i).

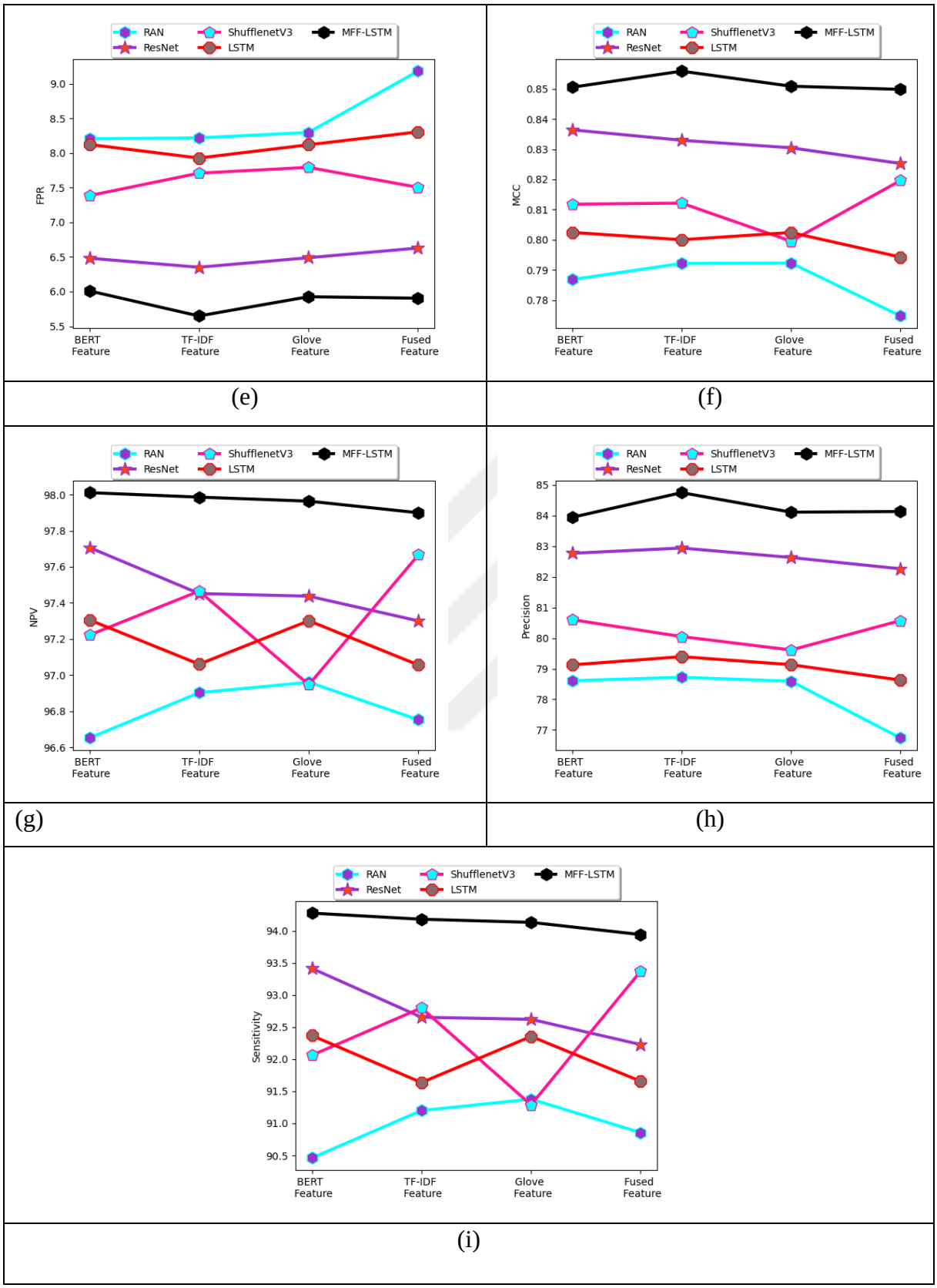

Figure 4.4: Comparative Evaluation of Recommended Fake News Detection Model's Feature Extraction over Diverse Conventional Classifier for Dataset 1 Concerning " (a) Accuracy, (b) F1-score, (c) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NPV, (h) Precision, (i)"Figure Continoued".

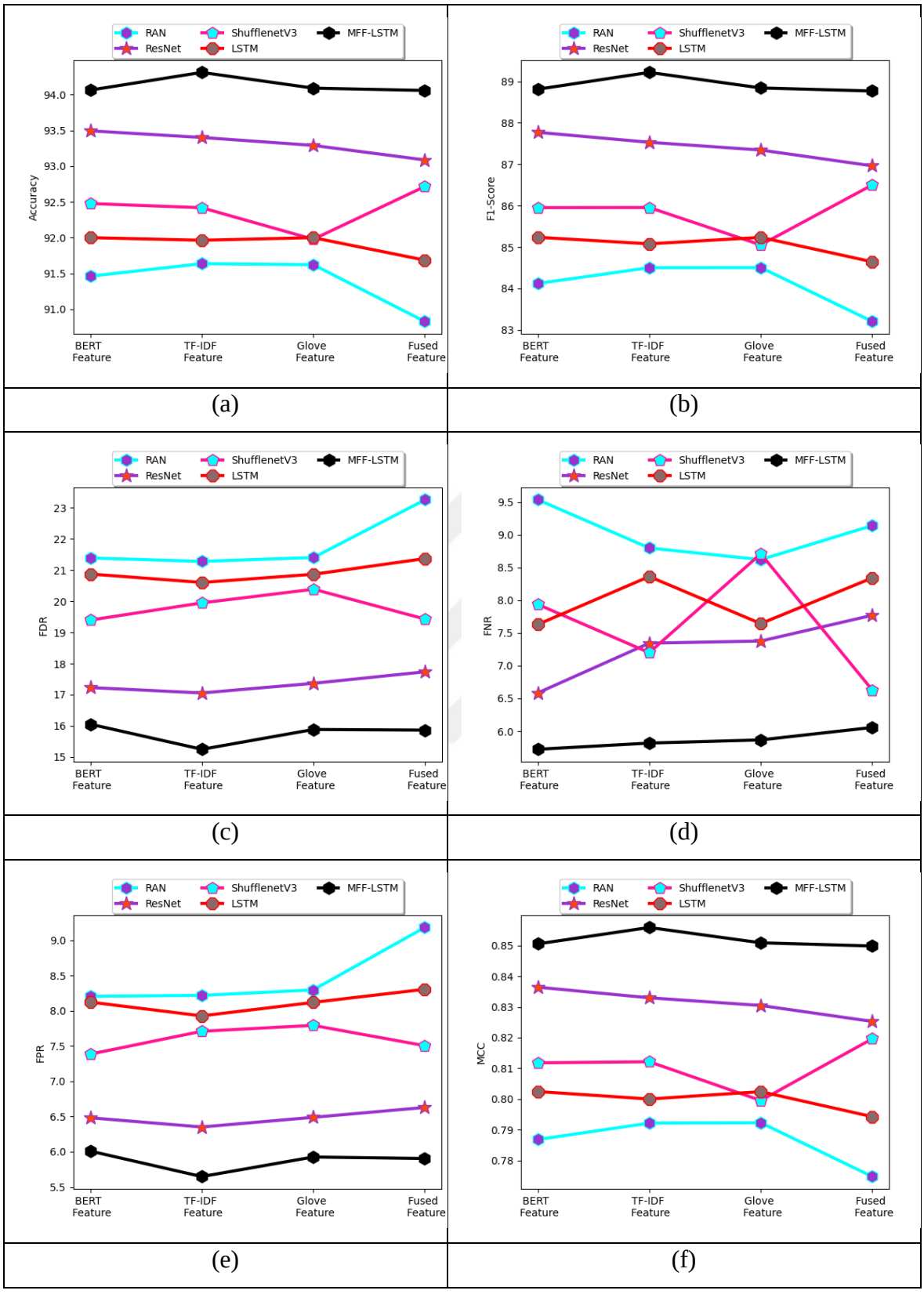

Figure 4.5: Comparative Evaluation of the Suggested Fake News Detection Model's Feature Extraction over Diverse Traditional Classifier for Dataset 2 Concerning " (a) Accuracy, (b) F1-score, (c) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NPV, (h) Precision, (i) Sensitiv**.**

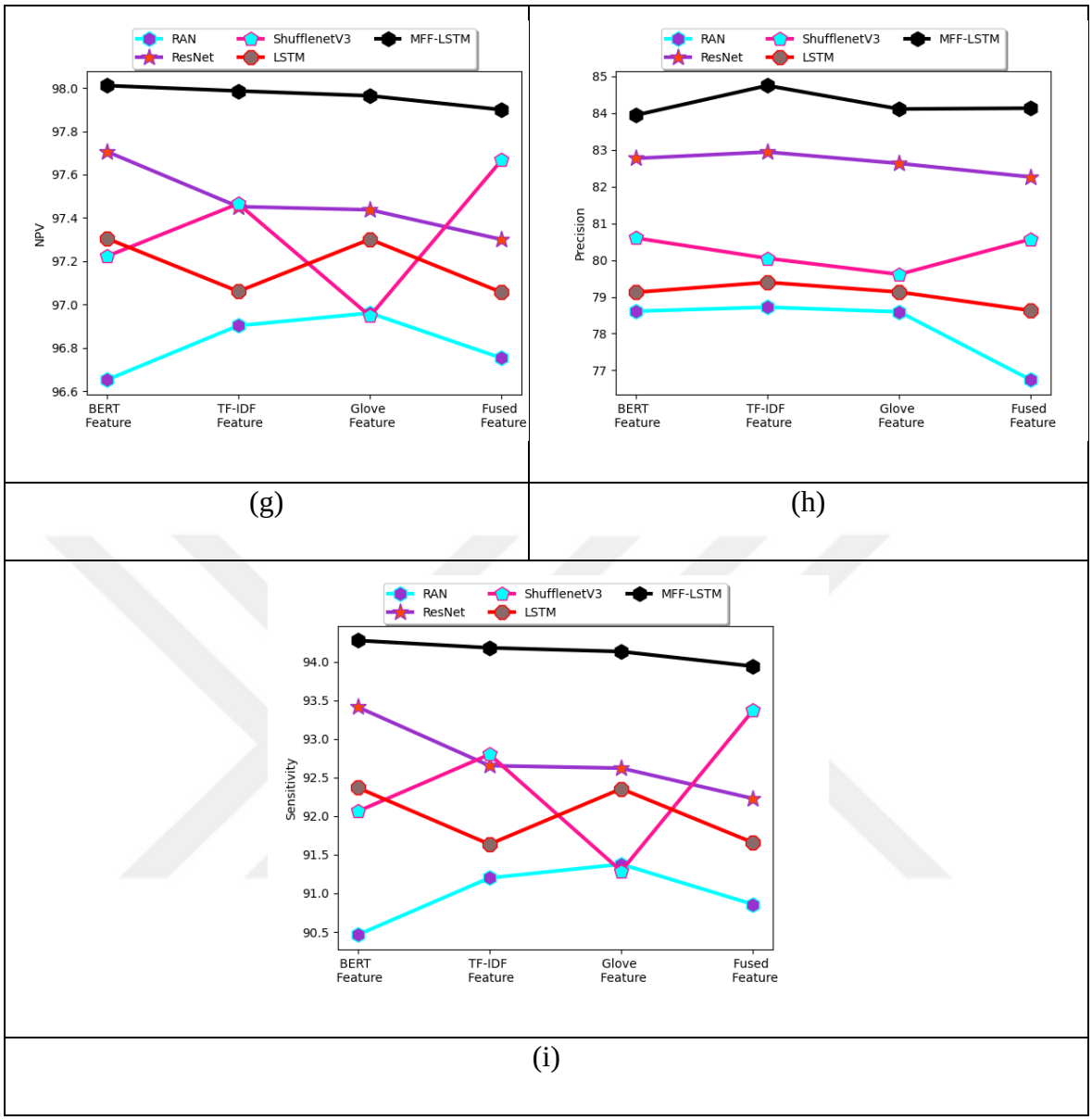

Figure 4.5: Comparative Evaluation of the Suggested Fake News Detection Model's Feature Extraction over Diverse Traditional Classifier for Dataset 2 Concerning " (a) Accuracy, (b) F1-score, (c) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NPV, (h) Precision, (i) Sensitiv " Figure Continued".

4.5 COMPARATIVE ANALYSIS OF LEARNING PERCENTAGE OVER TRADITIONAL CLASSIFIER IN DATASET 1 AND DATASET 2

The efficiency of the developed fake news detection model was evaluated across various traditional classifiers in dataset 1 and dataset 2, as depicted in Figure 12 and Figure 13. The assessment considered different learning percentage values for the presented work. In Figure

12 (h), when the learning percentage value is set to 55, the suggested fake news detection model exhibited enhanced precision is compared to RAN by 56%, ResNet by 24%, ShufflenetV3 by 21%, and LSTM by 67%, respectively. This underscored the superior quality of the designed fake news detection model compared to other conventional detection models.

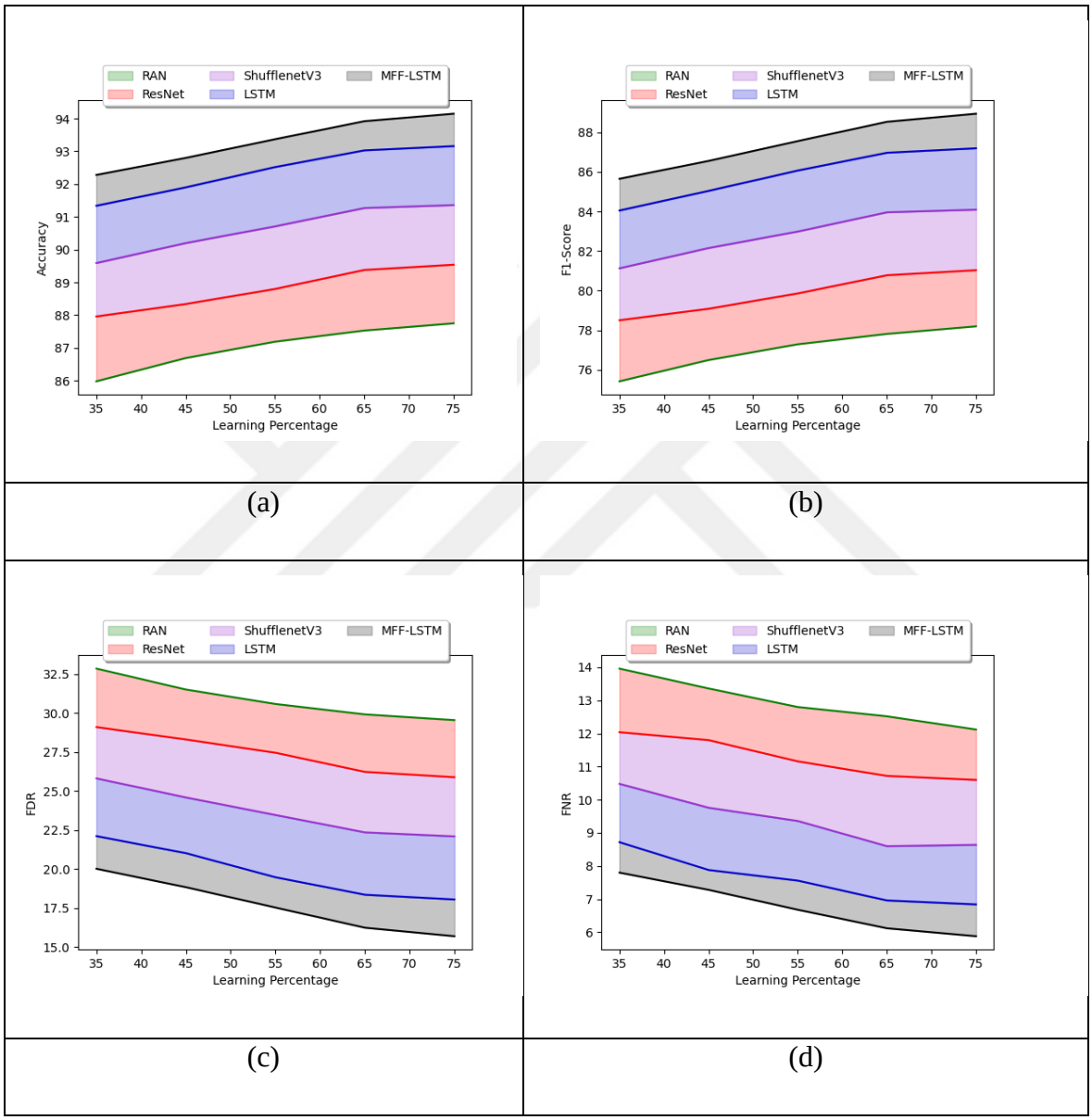

Figure 4.6: The Learning Percentage Revaluation of the Suggested Fake News Detection Model's Feature Extraction over Diverse Traditional Classifier for Dataset 1 Concerning " (a) Accuracy, (b) F1-score, (c) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NPV, (h) Precision, (i) Sensitivity".

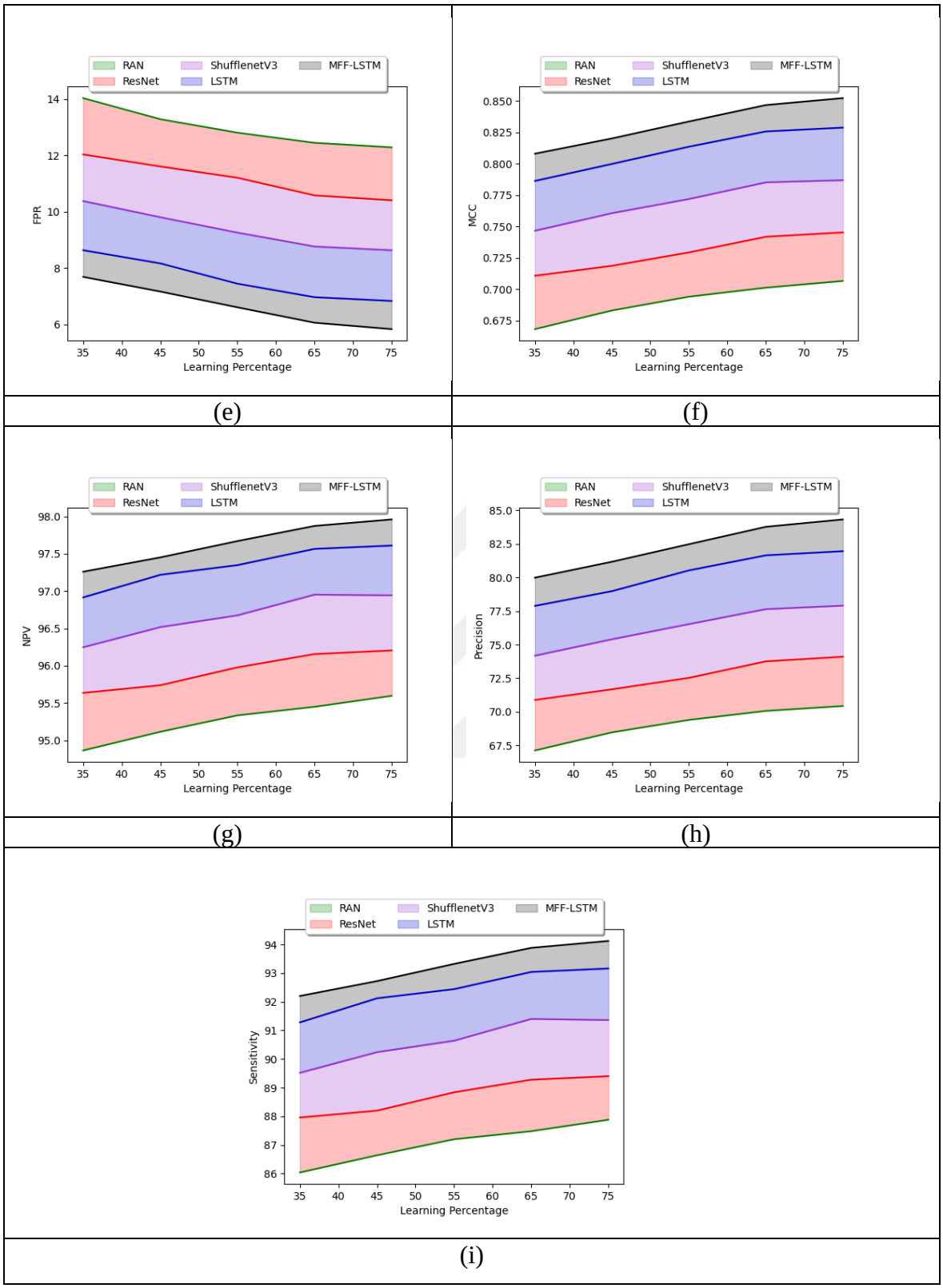

Figure 4.6: The Learning Percentage Revaluation of the Suggested Fake News Detection Model's Feature Extraction over Diverse Traditional Classifier for Dataset 1 Concerning " (a) Accuracy, (b) F1-score, (c) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NPV, (h) Precision, (i) Sensitivity". "Figure Continued".

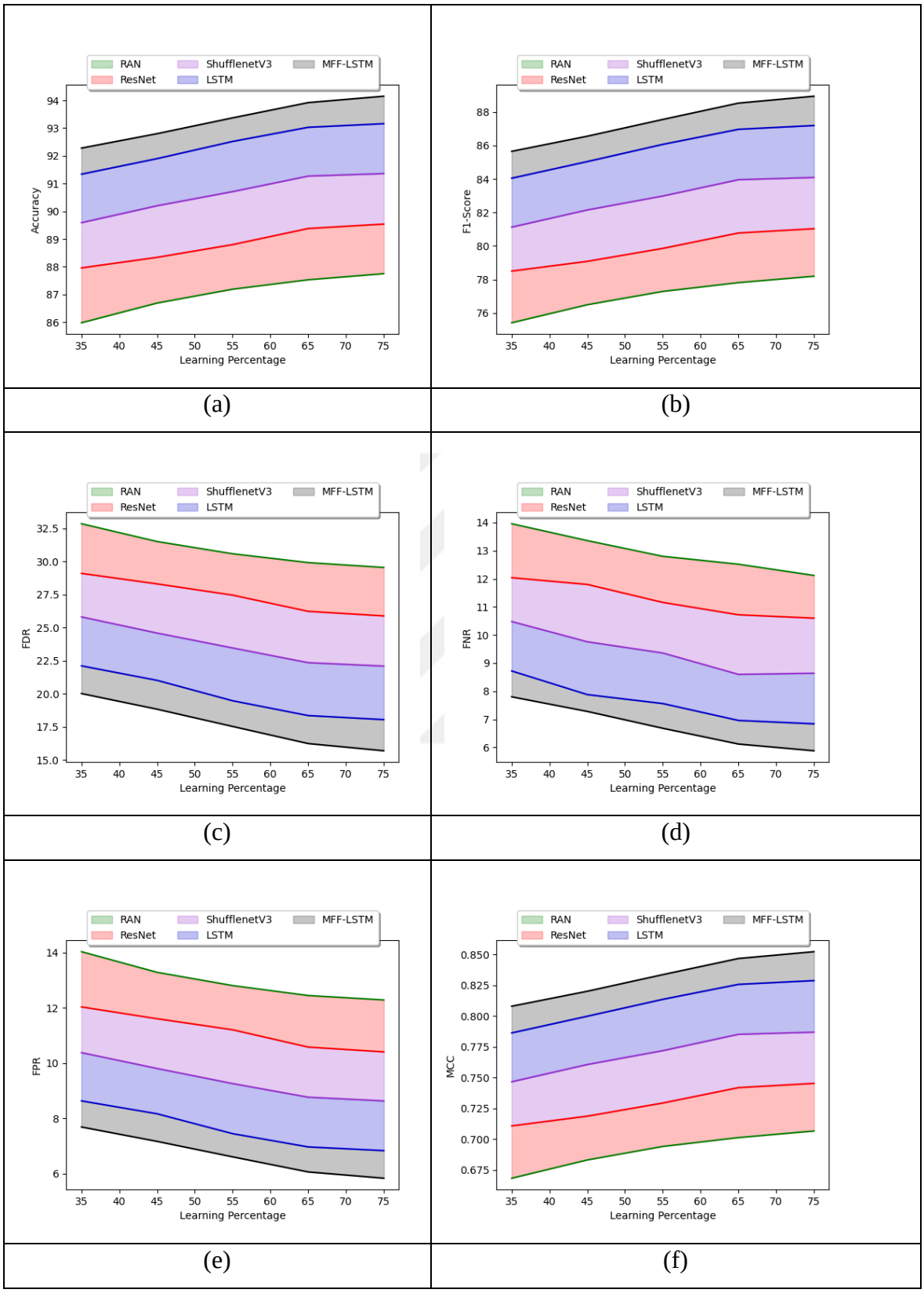

Figure 4.7: The Learning Percentage Evaluation of The Suggested Fake News Detection Model's Feature Extraction Over Diverse Traditional Classifier for Dataset 2 Concerning " (a) Accuracy, (b) F1-Score, (c) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NPV, (h) Precision, (i) Sensitivity".

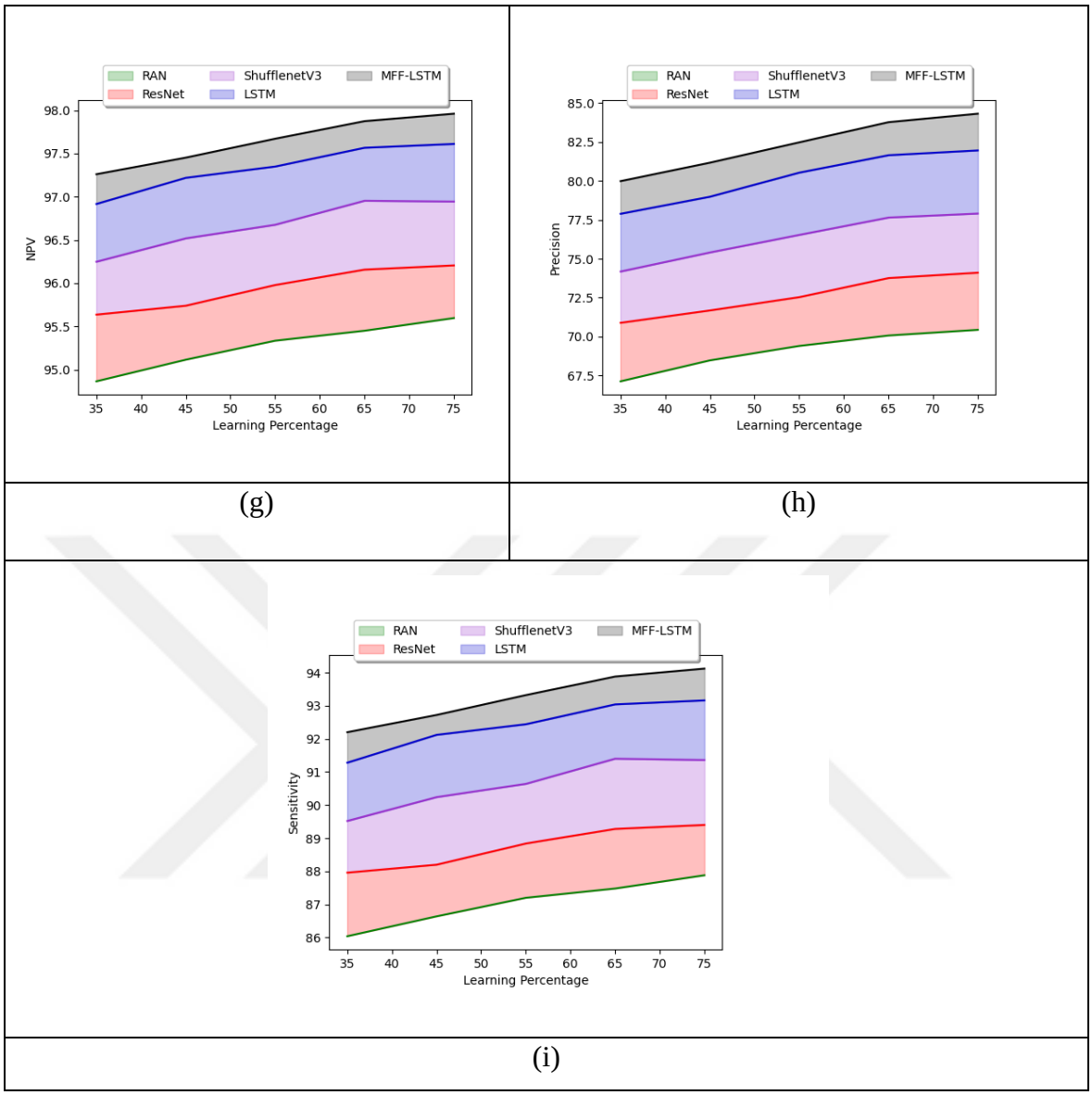

Figure 4.7: The Learning Percentage Evaluation of The Suggested Fake News Detection Model's Feature Extraction Over Diverse Traditional Classifier for Dataset 2 Concerning " (a) Accuracy, (b) F1-Score, (c) FDR, (d) FNR, (e) FPR, (f) MCC, (g) NPV, (h) Precision, (i) Sensitivity"." Figure Continued".

4.5 OVERALL PERFORMANCE ANALYSIS OF THE PROPOSED FAKE NEWS DETECTION MODEL FOR DATASET 1 AND DATASET 2

The comprehensive performance evaluation of the devised model of fake news detection across various traditional classifiers and algorithms in Dataset 1 and Dataset 2 is presented in Tables 2 and 3. The precision of the recommended fake news detection mechanism surpassed that of RAN by 78%, ResNet by 34%, ShufflenetV3 by 34%, and LSTM by 78%,

respectively. These findings highlight the superior functionality of the suggested fake news detection approach, indicating the highest precision rates.

Table 4.1: Overall Performance Analysis of the Proposed Fake News Detection Model for Dataset.

Classifier	RAN [26]	ResNet[27]	ShufflenetV3[28]	LSTM [29]	MFF-LSTM
“Accuracy”	89.25798	91.09151	92.92504	94.80409	95.73761
“Sensitivity”	89.27399	91.11703	92.92604	94.79108	95.73761
“Specificity”	89.25265	91.08301	92.9247	94.80843	95.73761
“Precision”	73.46683	77.30429	81.40558	85.88809	88.21729
“FPR”	10.74735	8.916994	7.075295	5.191574	4.262387
“FNR”	10.72601	8.882974	7.073961	5.208917	4.262387
“NPV”	96.14843	96.85148	97.52527	98.20155	98.53765
“FDR”	26.53317	22.69571	18.59442	14.11191	11.78271
“F1-score”	80.60273	83.64425	86.78515	90.12024	91.82373
“MCC”	73.93682	78.07437	82.31812	86.80087	89.08382

Table 4.2: Overall Performance Analysis of the Proposed Fake News Detection Model for Dataset.

Classifier	RAN [26]	ResNet[27]	ShufflenetV3[28]	LSTM[29]	MFF-LSTM
“Accuracy”	87.53431	89.29908	91.05138	92.94369	93.83647
“Sensitivity”	87.55188	89.24972	91.05266	92.94184	93.85275
“Specificity”	87.5157	89.35138	91.05002	92.94564	93.81923
“Precision”	88.13619	89.87706	91.5089	93.31403	94.14707
“FPR”	12.4843	10.64862	8.949983	7.054357	6.18077
“FNR”	12.44812	10.75028	8.94734	7.058158	6.147254
“NPV”	86.90537	88.69555	90.57166	92.55458	93.50956
“FDR”	11.86381	10.12294	8.491103	6.685968	5.852933
“F1-score”	87.84306	89.56229	91.28021	93.12757	93.99968
“MCC”	75.05457	78.58685	82.09162	85.87805	87.6643

5. CONCLUSION

In the pursuit of creating a cutting-edge adaptive learning model for fake news detection, several pivotal stages were undertaken to ensure a robust and effective system. The journey commenced with the collection of input text from benchmark datasets, laying the foundation for a comprehensive analysis. Rigorous pre-processing of the gathered text ensued, involving tasks such as the elimination of stop words, stemming, lemmatization, and adept handling of special characters. This preparatory phase aimed to refine the textual data, paving the way for subsequent in-depth analysis. Distinguishing itself from conventional approaches, the proposed model incorporated three distinct feature extraction techniques: BERT, TF-IDF, and GloVe Embedding. Each technique contributed a unique set of features, forming a comprehensive amalgamation that endowed the model with the capability to discern intricate patterns and features crucial for fake news detection. This holistic approach was a strategic move, capitalizing on the individual strengths of each method to enhance the overall efficacy of the model. The fusion of these features was meticulously executed through the MFF-LSTM model. This innovative integration allowed for a dynamic and multiscale application of the extracted features across the layers of the LSTM network. This sophisticated architecture aimed to elevate the model's performance, providing it with a nuanced understanding of the input text and significantly bolstering its ability to identify fake news. To ascertain the model's prowess, a critical stage involved the comparison of its performance against traditional approaches to fake news detection. This comparative analysis was deemed indispensable for gauging the effectiveness and superiority of the developed system. The ultimate aspiration is to cultivate a model that not only achieves high accuracy but also acts as a formidable deterrent against the dissemination of false information and misinformation, thereby contributing to the safeguarding of reliable information channels. The model's performance is contingent on the quality and diversity of the training data. when the learning percentage value is set to 60, the suggested fake news detection model exhibited enhanced accuracy is compared to RAN by 34%, ResNet by 44%, ShufflenetV3 by 31%, and LSTM by 27%, respectively. This underscored the superior quality of the designed fake news detection model compared to other conventional detection models. If the training dataset contains biases or lacks representation of certain types of fake news, the model may struggle to generalize effectively. While the proposed fake news detection model showcases significant advancements, ongoing efforts in addressing its limitations and exploring future

avenues are crucial for ensuring its relevance and effectiveness in an ever-evolving landscape of misinformation.

REFERENCES

[1] Jamal Abdul Nasir, Osama Subhani Khan, Iraklis Varlamis, "Fake news detection: A hybrid CNN-RNN based deep learning approach", Elsevier International Journal of Information Management Data Insights, Vol. 1, Issue 1, pp. 100007, April 2021.

[2] T. Jiang, J. P. Li, A. U. Haq, A. Saboor and A. Ali, "A Novel Stacking Approach for Accurate Detection of Fake News," in IEEE Access, vol. 9, pp. 22626-22639, 2021.

[3] H. Saleh, A. Alharbi and S. H. Alsamhi, "OPCNN-FAKE: Optimized Convolutional Neural Network for Fake News Detection," in IEEE Access, vol. 9, pp. 129471-129489, 2021.

[4] M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi and B. -W. On, "Fake News Stance Detection Using Deep Learning Architecture (CNN-LSTM)," in IEEE Access, vol. 8, pp. 156695-156706, 2020.

[5] T. Huu Do, M. Berneman, J. Patro, G. Bekoulis and N. Deligiannis, "Context-Aware Deep Markov Random Fields for Fake News Detection," in IEEE Access, vol. 9, pp. 130042-130054, 2021.

[6] L. Ying, H. Yu, J. Wang, Y. Ji and S. Qian, "Fake News Detection via Multi-Modal Topic Memory Network," in IEEE Access, vol. 9, pp. 132818-132829, 2021.

[7] Liyakathunisa Syed, Abdullah Alsaeedi, Lina A. Alhuri, Hutaf R. Aljohani, "Hybrid weakly supervised learning with deep learning technique for detection of fake news from cyber propaganda", Elsevier Array, Vol.19, pp. 100309, September 2023.

[8] Balasubramanian Palani, Sivasankar Elango, "BBC-FND: An ensemble of deep learning framework for textual fake news detection", Elsevier Computers and Electrical Engineering, Vol. 110, pp.108866, September 2023.

[9] M. Kang, J. Seo, C. Park and H. Lim, "Utilization Strategy of User Engagements in Korean Fake News Detection," in IEEE Access, vol. 10, pp. 79516-79525, 2022.

[10] H. Ali et al., "All Your Fake Detector are Belong to Us: Evaluating Adversarial Robustness of Fake-News Detectors Under Black-Box Settings," in IEEE Access, vol. 9, pp. 81678-81692, 2021.

[11] M. Narra et al., "Selective Feature Sets Based Fake News Detection for COVID-19 to Manage Infodemic," in IEEE Access, vol. 10, pp. 98724-98736, 2022.

[12] P. Wei, F. Wu, Y. Sun, H. Zhou and X. -Y. Jing, "Modality and Event Adversarial Networks for Multi-Modal Fake News Detection," in IEEE Signal Processing Letters, vol. 29, pp. 1382-1386, 2022.

[13] Qin Zhang, Zhiwei Guo, Yanyan Zhu, Pandi Vijayakumar, Aniello Castiglione, Brij B. Gupta, "A Deep Learning-based Fast Fake News Detection Model for Cyber-Physical Social Services", ElsevierPattern Recognition Letters, Vol. 168, pp. 31-38, April 2023.

[14] Olusoji B. Okunoye, Ayei E. Ibor, "Hybrid fake news detection technique with genetic search and deep learning", Elsevier Computers and Electrical Engineering, Vol. 103, pp.108344, October 2022.

[15] Azka Kishwar, Adeel Zafar, "Fake news detection on Pakistani news using machine learning and deep learning", Elsevier Expert Systems with Applications, Vol. 211, pp.118558, January 2023.

[16] Tavishee Chauhan M.E, Hemant Palivela PhD, "Optimization and improvement of fake news detection using deep learning approaches for societal benefit", Elsevier International Journal of Information Management Data Insights, Vol. 1, Issue 2, pp. 100051, November 2021.

[17] I. Kadek Sastrawan, I.P.A. Bayupati, Dewa Made Sri Arsa, "Detection of fake news using deep learning CNN–RNN based methods", Elsevier ICT Express, Vol. 8, Issue 3, Pages 396-408, September 2022.

[18] Somya Ranjan Sahoo, B.B. Gupta, "Multiple features based approach for automatic fake news detection on social networks using deep learning", Elsevier Applied Soft Computing, Vol. 100, pp. 106983, March 2021.

[19] Mohammadreza Samadi, Maryam Mousavian, Saeedeh Momtazi, "Deep contextualized text representation and learning for fake news detection", Elsevier Information Processing & Management, Vol. 58, Issue 6, 102723, November 2021.

[20] Eduri Raja, Badal Soni, Samir Kumar Borgohain, "Fake news detection in Dravidian languages using transfer learning with adaptive finetuning", Elsevier Engineering Applications of Artificial Intelligence, Vol. 126, Part A, pp. 106877, November 2023.

[21] Haosen Wang, Pan Tang, Hanyue Kong, Yilun Jin, Chunqi Wu, Linghong Zhou, "DHCF: Dual disentangled-view hierarchical contrastive learning for fake news detection on social media", Elsevier Information Sciences, Vol. 645, pp. 119323, October 2023.

[22] Anshika Choudhary, Anuja Arora, "Assessment of bidirectional transformer encoder model and attention based bidirectional LSTM language models for fake news detection", Elsevier Journal of Retailing and Consumer Services, Vol. 76, pp. 103545, January 2024.

[23] Jiaheng Hua, Xiaodong Cui, Xianghua Li, Keke Tang, Peican Zhu, "Multimodal fake news detection through data augmentation-based contrastive learning", Elsevier Applied Soft Computing, Vol. 136, pp. 110125, March 2023.

[24] Marcos Paulo Silva Gôlo, Mariana Caravanti de Souza, Rafael Geraldeli Rossi, Solange Oliveira Rezende, Bruno Magalhães Nogueira, Ricardo Marcondes Marcacini, "One-class learning for fake news detection through multimodal variational autoencoders", Elsevier Engineering Applications of Artificial Intelligence, Vol. 122, pp. 106088, June 2023.

[25] Marjan Hosseini, Alireza Javadian Sabet, Suining He, Derek Aguiar, "Interpretable fake news detection with topic and deep variational models", Esevier Online Social Networks and Media, Vol. 36, 100249, July 2023.

[26] X. Yang, M. Zhang, W. Li and R. Tao, "Visible-Assisted Infrared Image Super-Resolution Based on Spatial Attention Residual Network," IEEE Geosciences' and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 7002705.

[27] A. Krueangsai and S. Supratid, "Effects of Shortcut-Level Amount in Lightweight ResNet of ResNet on Object Recognition with Distinct Number of Categories," 2022 International Electrical Engineering Congress (iEECON), Khon Kaen, Thailand, 2022, pp. 1-4.

[28] BAIRABOINA SAI SAMBASIVA RAO AND BATTULA SRINIVASA RAO"An Effective WBC Segmentation and Classification Using MobilenetV3–ShufflenetV2 Based Deep Learning Framework"School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh 522237, India.

[29] B. Majumdar, M. RafiuzzamanBhuiyan, M. A. Hasan, M. S. Islam and S. R. H. Noori, "Multi Class Fake News Detection using LSTM Approach," 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART), MORADABAD, India, 2021, pp. 75-79.

[30] Aswini ,Priyanka Tilak ,Simrat Ahluwalia ,Nibrat Lohia "Fake News Detection: A Deep Learning Approach"SMU Data Science Review.

[31] Z. Jianqiang and G. Xiaolin, "Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis,"IEEE Access, vol. 5, pp. 2870-2879, 2017.

[32] K. S. Srujan(B), S. S. Nikhil, H. Raghav Rao, K. Karthik, B. S. Harish and H. M. Keerthi Kumar"Classification of Amazon Book Reviews Based on Sentiment Analysis"Department of Information Science and Engineering, Sri Jayachamarajendra College of Engineering, Mysuru 570006, Karnataka, India.

[33] Danasingh, A.A.G.S., Subramanian, A. & Epiphany, J. "Identifying redundant features using unsupervised learning for high-dimensional data". SN Applied Sciences Vol.2, 1367, 2020.

[34] Jaideepsinh K. Raulji and Jatinderkumar R. Saini,"Stop-Word Removal Algorithm and its Implementation for Sanskrit Language"International Journal of Computer Applications (0975 – 8887) Volume 150 – No.2, September 2016.

[35] Abdullah Wahbeh,Mohammed Al-Kabi, Qasem Al-Radaideh, Emad Al-Shawakfa and Izzat Alsmadi"The Effect of Stemming on Arabic Text Classification: An Empirical Study" International Journal of Information Retrieval Research, 1(3), 54-70, July-September 2011.

[36] Z. Ba and B. Hu, "A Universal Bert-Based Front-End Model for Mandarin Text-To-Speech Synthesis," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 6074-607.

[37] Zhang Y, Zhou Y, Yao J." Feature Extraction with TF-IDF and Game-Theoretic Shadowed Sets". Information Processing and Management of Uncertainty in Knowledge-Based Systems. 2020 May 18; 1237:722–33.

[38] Laura Dipietro, Angelo M. Sabatini, Senior Membe and Paolo Dario" A Survey of Glove-Based Systemsand Their Applications "IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, VOL. 38, NO. 4, JULY 2008.

[39] X. Zhao, H. Zhang, Y. Zhang, T. K. Sarkar and S. -W. Ting, "Efficient Modelling of Multiscale Structures Using Higher-Order Method of Moments," IEEE Journal on Multiscale and Multiphysics Computational Techniques, vol. 2, pp. 78-83, 2017.

[40] M. Park and S. Chai, "Constructing a User-Centered Fake News Detection Model by Using Classification Algorithms in Machine Learning Techniques," IEEE Access, vol. 11, pp. 71517-71527, 2023.

[41] N. Seddari, A. Derhab, M. Belaoued, W. Halboob, J. Al-Muhtadi and A. Bouras, "A Hybrid Linguistic and Knowledge-Based Analysis Approach for Fake News Detection on Social Media," IEEE Access, vol. 10, pp. 62097-62109, 2022.

[42] T. Mahara, V. L. H. Josephine, R. Srinivasan, P. Prakash, A. D. Algarni and O. P. Verma, "Deep vs. Shallow: A Comparative Study of Machine Learning and Deep Learning Approaches for Fake Health News Detection," IEEE Access, vol. 11, pp. 79330-79340, 2023.

[43] A. A. Obaid, H. Khotanlou, M. Mansoorizadeh and D. Zabihzadeh, "Robust Semi-Supervised Fake News Recognition by Effective Augmentations and Ensemble of Diverse Deep Learners,"IEEE Access, vol. 11, pp. 54526-54543, 2023.

[44] .J. Chen, C. Jia, H. Zheng, R. Chen and C. Fu, "Is Multi-Modal Necessarily Better? Robustness Evaluation of Multi-Modal Fake News Detection," IEEE Transactions on Network Science and Engineering, vol. 10, no. 6, pp. 3144-3158, Nov.-Dec. 2023.

[45] H. Choi and Y. Ko, "Using Adversarial Learning and Biterm Topic Model for an Effective Fake News Video Detection System on Heterogeneous Topics and Short Texts," in IEEE Access, vol. 9, pp. 164846-164853, 2021.

[46] C. -O. Truică, E. -S. Apostol, R. -C. Nicolescu and P. Karras, "MCWDST: A Minimum-Cost Weighted Directed Spanning Tree Algorithm for Real-Time Fake News Mitigation in Social Media," IEEE Access, vol. 11, pp. 125861-125873, 2023.

[47] Ravish, R. Katarya, D. Dahiya and S. Checker, "Fake News Detection System Using Featured-Based Optimized MSVM Classification," IEEE Access, vol. 10, pp. 113184-113199, 2022.

[48] G. Shan, B. Zhao, J. R. Clavin, H. Zhang and S. Duan, "Poligraph: Intrusion-Tolerant and Distributed Fake News Detection System," IEEE Transactions on Information Forensics and Security, vol. 17, pp. 28-41, 2022.

[49] P. Li, X. Sun, H. Yu, Y. Tian, F. Yao and G. Xu, "Entity-Oriented Multi-Modal Alignment and Fusion Network for Fake News Detection," IEEE Transactions on Multimedia, vol. 24, pp. 3455-3468, 2022.

[50] M. Almohammed, H. M. Farhan, R. A. S. Naseri and A. K. Turkben, "Data Mining And Analysis For Predicting Electrical Energy Consumption," *2022 International Conference on Artificial Intelligence of Things (ICAIoT)*, Istanbul, Turkey, 2022, pp. 1-7, doi: 10.1109/ICAIoT57170.2022.10121820

[51] A. R. Al-Aloosi, H. M. Farhan, R. A. S. Naseri, A. K. Turkben, A. K. Mustafa and M. G. F. Al-Obadi, "Face Recognition System Using Local Binary Pattern with Binary Dragonfly Algorithm to Feature Selection," *2022 International Conference on Artificial Intelligence of Things (ICAIoT)*, Istanbul, Turkey, 2022, pp. 1-10, doi: 10.1109/ICAIoT57170.2022.10121837.

[52] M. G. Al-Obadi, H. M. Farhan, R. A. S. Naseri, A. K. Turkben, A. K. Mustafa and A. R. Al-Aloosi, "Data Mining Techniques for Extraction and Analysis of Covid-19 Data," *2022 International Conference on Artificial Intelligence of Things (ICAIoT)*, Istanbul, Turkey, 2022, pp. 1-7, doi: 10.1109/ICAIoT57170.2022.10121870.

[53] D. A. Khalef, A. K. Turkben, H. M. Farhan and R. A. S. Naseri, "Optic Disc Segmentation In Human Retina Images With Meta Heuristic Optimization," *2022 International Conference on Artificial Intelligence of Things (ICAIoT)*, Istanbul, Turkey, 2022, pp. 1-6, doi: 10.1109/ICAIoT57170.2022.10121896.

[54] Naseri, Raghda Awad Shaban, Ayça Kurnaz, and Hameed Mutlag Farhan. "Optimized face detector-based intelligent face mask detection model in IoT using deep learning approach." Applied Soft Computing 134 (2023): 109933.