RESEARCH ARTICLE

Big Data ML-Based Fake News Detection Using Distributed Learning

ALAA ALTHENEYAN[ID] **AND ASEEL ALHADLAQ**[ID]

Department of Computer Science and Engineering, College of Applied Studies and Community Services, King Saud University, Riyadh 11495, Saudi Arabia

Corresponding author: Alaa Altheneyan (atheneyan@ksu.edu.sa)

ABSTRACT Users rely heavily on social media to consume and share news, facilitating the mass dissemination of genuine and fake stories. The proliferation of misinformation on various social media platforms has serious consequences for society. The inability to differentiate between the several forms of false news on Twitter is a major obstacle to effective detection of fake news. Researchers have made progress toward a solution by emphasizing methods for identifying fake news. The dataset FNC-1, which includes four categories for identifying false news, will be used in this study. The state-of-the-art methods for spotting fake news are evaluated and compared using big data technology (Spark) and machine learning. The methodology of this study employed a decentralized Spark cluster to create a stacked ensemble model. Following feature extraction using N-grams, Hashing TF-IDF, and count vectorizer, we used the proposed stacked ensemble classification model. The results show that the suggested model has a superior classification performance of 92.45% in the F1 score compared to the 83.10 % F1 score of the baseline approach. The proposed model achieved an additional 9.35% F1 score compared to the state-of-the-art techniques.

INDEX TERMS Big data, machine learning, fake news, ensemble learning, social media.

I. INTRODUCTION

The use of social media platforms to disseminate and digest media has increased in recent years. Social networking sites like Facebook and Twitter generate daily data [1]. It is no secret that the internet is a goldmine of information, especially recent news [2]. The proliferation of fake news is directly attributable to the internet's user-friendly nature. Since fake news is often presented as factual, it is often shared on social media. Often, this data is spread for profit or influencing politics. The effects of fake news on society as a whole are profound. In the light of its profound impacts, fixing this issue is crucial [3]. Multiple instances of false news were reported to have spread on social media during the 2016 US elections, including the presidential election and the nomination of a new Air Marshal in India [4]. The dissemination of false information has negatively affected people's mental health and society as a whole [5].

The associate editor coordinating the review of this manuscript and approving it for publication was Chong Leong Gan[ID].

Many automatically assume that the news is either bogus or legitimate based on the article's content. Techniques based on news content use methods for collecting data and tone from fake news stories. The goal of style-based methods for de-detecting false news is to utilize the manipulators' writing styles for detection. By examining certain language features, we can distinguish fake news from the real thing [3]. However, false news is created with the intent of fooling readers. Thus, improving the detection of false news using news content style is a difficult problem. To assist in avoiding the difficult and time-consuming human work of fact-checking, the Natural Language Processing (NLP) industry has shown considerable interest in automatic recognition of fake news [6], [7]. Determining the integrity of news is a difficult task, even for automated approaches [8]. Familiarizing with what other news outlets say on the same issue might be a useful starting point for recognizing false news. Identifying a person's position is the purpose of this phase. Multiple tasks, such as evaluating online arguments [9], [10], verifying the integrity of Twitter rumors [11], [12],

Headline	Seven girls fall pregnant after five-days school trip in Bosnia and Herzegovina.
Agree	Seven girls, aged 13 to 15, have fallen pregnant after a five-days school trip to their country's capital city and their parents are being blamed [..].
Disagree	On December 17, a site called In Serbia.info published a 400-word piece under the sensational headline: BiH: Seven Primary School Students Pregnant After Five-Days Excursion[..].
Discuss	School field trips aren't what they used to be. Just ask the furious parents of seven teenage girls who became pregnant on a five-days school trip to Sarajevo, the capital city of Bosnia[..].
Unrelated	Lebanese authorities are holding a daughter and an ex-wife of the head of the ISIS jihadist group, Abu Bakr al-Baghdadi, the interior minister said. It was initially reported that a wife and son of the self-proclaimed had been arrested in November[..].

FIGURE 1. Overview of the headline and text bodies with their respective stances.

or understanding the argumentation structure of seminal works [13], [14], have traditionally relied on position identification.

In the first example of evaluating the first False News Challenge (FNC-1), a media news source discusses a topic to create automated fake news detection systems using AI technology and machine learning. Almost fifty groups from industry and academics worked on this problem. One of the objectives of the FNC-1 challenge is to track out a media production dealing with a certain title. It might support, challenge, or have nothing to do with the title. Four potential vantage points from which an essay is to be written. The guidelines, dataset, and grading criteria for the FNC-1 challenge are all available on their site. These topics are further shown in Figure 1, which depicts the results of four distinct research.

Multiple deep learning and Recurrent Neural Networks (RNN), as well as their modifications, including Convolution Neural Networks (CNN) [15], are often employed for NLP tasks and have shown to perform magnificently on NLP-related tasks [16], [17], [18].

A. OVERVIEW OF FAKE NEWS DETECTION

In 2017, Facebook released a white paper that explored the risks of online communication and the management of being one of the most prominent social media platforms today. Weedon, Nuland, and Stamos also noticed the growing challenge of using the enigmatic phrase "fake news," and proclaimed that "the overuse and misapplication of the term "fake news" might be challenging since we cannot understand or adequately address these concerns without shared definitions" [19]. The word can apply to anything from virtually incorrect news articles to deceptions, April

Fools' jokes, rumors, clickbait, or stated opinions posted online with incorrect facts.

In this research work, "fake news" is defined as a written article that is manifestly untrue and falsely disseminated without being authentic mostly accompanird by malicious intents. This definition includes three important textual, visual, and audio bases. Other elements such as video-based fake news and audio, are typically ignored when referring to textual fake news; additionally, each element has its linguistic complexities that necessitate different machine learning and deep learning algorithms to detect and solve problems such as 'Deep Fake,' etc. The notion also implies that fake news might be fact-checked, an important characteristic. Therefore, the claims may be checked to see if they are true or false. Because rumors are usually hard to verify, they are deleted from the definition because of this inclusion. Conspiracy theories are classed as rumors because they are persistent rumors that are difficult to refute. False information concerning the entertainment sector, including hoaxes and April Fools' gags, is not permitted because the objective must be harmful. Furthermore, the goal is infamous as it seeks to affect public opinion in favor of a specific message. It also removes text bits that were mistakenly published improperly, such as transposed numbers.

A model of the connection between headlines and news content is necessary for identifying clickbait. It is also crucial to tell the difference between false news and clickbait. The term "clickbait" refers to articles with enticing headlines written to attract online audience or traffic; when people click on such a headline, they end up at a different website with poorly written articles that have nothing to do with the subject line. So, clickbait is written with one goal: getting more people to visit a website that relies on advertising to make

money. The motive is monetary gain rather than furthering a political agenda via disseminating false information.

A great example is the deliberately spread of false news about Hillary Clinton by Russian trolls in the 2016 presidential election campaign, which was designed to affect people's voting choices away from Hillary and toward Donald Trump. This instance demonstrates how dangerous it can be when false information spreads on critical issues. Of course, there's another problem with false news, toxic information is spread for no reason to sow doubt, stir up chaos, and make it difficult for readers to tell fact from fiction.

1) SOCIAL MEDIA AND FAKE NEWS

Global knowledge dissemination has been democratized because of technological advancements and the emergence of social media. Important news organizations have invested heavily in digital journalism, generating content for media platforms, and growing their reach via social media and online tools. Furthermore, online social media platforms are becoming most important sites for information spreading. Dissemination of information allows for the exchange of ideas and the connectivity of previously inaccessible locations. It enables users to generate opinions about the information platforms offer from many perspectives.

In the past, media companies have invested heavily in creating their presence online, with online media networking sites playing a significant role. They use social media platforms such as Facebook and Twitter to promote their material, spread information/news, and develop a network of individuals they may engage with. On the other hand, users benefit from social media's technical developments since people now have access to a wide range of information sources.

The current digital landscape for information dissemination and the challenges that media organizations face in an ever-present media environment have resulted in substantial changes in how news organizations are founded. Economic, technical, and social pressures have combined with the desire to be always noticeable, race of reporting with similar speed and excitement, getting followers, creating an atmosphere where fake news is prevalent.

The latest technological advancements in social media have undoubtedly provided a hostile environment for spreading online lies in a primarily deregulated media financed and driven by advertising. The motivation for good is usually overshadowed by the desire for profit, which significantly influences how the medium changes over time. According to the above, fake news exists on social media alongside real news, and the difficulty appears to be distinguishing them. While fake news is not new, the speed it travels and the worldwide reach of the instruments that can distribute it are unprecedented. Consequently, fake news emerges on social media in the same context as actual news, and the problem appears to be discerning between the two. While fake news is not a new phenomenon, the pace and quantity with which it is distributed have changed: social media platforms

such as Twitter, Facebook, and Instagram provide an ideal ground for quickly transmitting fake news. Furthermore, bots are increasingly being utilized to distort information, disrupt social media conversations, and draw users' attention, according to the same author.

2) USERS' RE-SHARING BEHAVIOR AND FAKE NEWS

From the perspectives discussed so far, it can be deduced that social media sites play an essential role in disseminating false information. Furthermore, internet users are to blame for spreading false information. There are two main types of data sharing on online sharing sites: self-disclosure, in which a user voluntarily discloses private information, and re-sharing, in which a user distributes material already created by another user of the site or a third party. Distributing low-quality, erroneous, or purposefully misleading material may have negative implications, such as spreading false news, but spreading high-quality information can assist in development of a more informed community. One of the most common ways information is disseminated online is by re-sharing, which includes retweeting, re-posting, re-vining, and re-blogging. In social media, for instance, it is common practice for users to write articles, distribute them among their networks, and engage in related online discourse. Social media users may engage in this practice with various apps. Sharing information rapidly is essential in many situations, including political campaigns and times of crisis, and therefore sites like Twitter, YouTube, and Facebook have become more important. Individuals are also using social media accounts for news production and dissemination.

In the case of social media, for instance, someone may spread false information (or even create a fake tale and post it). Resharing is a feature of many social media sites, so if one person shares a story, it increases the likelihood that others will do the same. Several proposed remedies are present, but there is still much disagreement over what constitutes ''fake news'', how it spreads, and how it affects social and political outcomes. Multiple major actors—including social media platforms, users, and groups against the spread of fake news—may be able to control the spread of false information on the internet. This brief theoretical overview of the Uses and gratifications theory (UGT), the filter bubble phenomenon, and social media re-sharing behavior provides important context for the current investigation. According to UGT's research, the Ellinika-Hoaxes-Facebook demographic represents an engaged audience searching for high-quality news and information from sources outside their echo chamber via media consumption. Know that this demographic is engaged, actively looking for information and trying to confirm the integrity of rumours they may have seen on social media. Users' familiarity with the Internet, social media, and other media is crucial for identifying the prevalence of false news on these platforms and stopping its spread. To properly answer the Formulation of research question (RQ) and draw conclusions on how members of the Ellinika-Hoaxes-Facebook group use particular media to
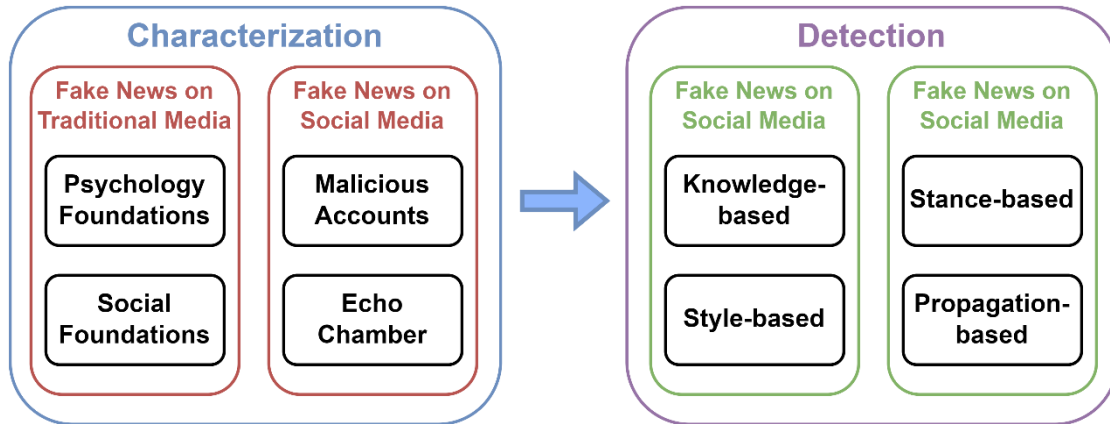
FIGURE 2. Category of fake news on social media.

detect and prevent the spread of false news, it is necessary to conduct research into their online behavior.

B. FAKE NEWS CHARACTERIZATION

The principle of fake news has two components: authenticity and purpose. The word "authenticity" refers to the fact that misleading news often contains false information that may be demonstrated to be untrue. Conspiracy theories, for example, are not included in the definition of fake news since it is nearly hard to tell whether they are real or false in most situations. According to the second component, the erroneous material's objective was to deceive the reader. Figure 2 represents the category of fake news on social media. The characterization module represents fake news belonging to traditional media and social media. The second module shows the fake news detection techniques used for both traditional and social media.

First, to identify fake news, understand the text context and the procedure to categorize it. It is vital to begin by characterization when developing detection models, and it is also necessary to grasp what fake news is before attempting to identify it. It is also not easy to develop a universally agreed definition for "fake news: Stories that are purposely and verifiably misleading and mislead readers". As per Wikipedia, deliberate misinformation or hoaxes spread via multiple online platforms and news channels or digital social media constitute a sort of fake journalism or propaganda [20]. Today's fake news is manipulative and diversified in topics, techniques, and platforms. It consists of two components: authenticity and intent. Fake news material that contains inaccuracies that may be verified falls under authenticity. However, it excludes conspiracy theories because they are difficult to prove actual or wrong in most circumstances. The second part refers to the misleading material written to deceive the reader.

C. TRADITIONAL MEDIA FAKE NEWS

The media ecosystem supporting the spread of false information has grown and evolved throughout time, including print,

broadcast, social media, and digital platforms. Before the rise of social media, this was seen as a concern because of its role in disseminating false information. Multiple psychological and social scientific foundations are used to characterize the effects of false news on individuals and the social knowledge environment. Humans aren't great at spotting believable stories from those who aren't. Several psychological and perceptual theories explain this phenomenon and the impact of misleading information. Traditional false news exploits readers' emotional vulnerabilities. Incorrect information is more likely to irritate clients due to the following two major factors:

- Customers with naive realism believe that their view of the world is valid and that others who disagree with them are irrational or dishonest [21].
- People are more likely to be presented with data that backs with their existing worldview. The cognitive biases that are part of the human condition led consumers to regularly confuse fake news with the genuine thing [22].

By analysing the news ecosystem as a whole, we may be able to pinpoint some of the societal factors that fuel the spread of disinformation. Theories of Social Identity [23] and Normative Influence [5] argue that the need for others' approval is central to a person's sense of self and identity, which increases the likelihood that users will prefer the anonymity and security of online platforms when obtaining and sharing news content, even if it is false.

D. THE EXTRACTION OF FEATURES

Unlike social media, where additional social data may help identify false news, conventional news organizations rely on content like text and photographs to spot and identify fake news. Some representative features of false news were shown in Figure 3. We will next examine how to extract and disseminate relevant data from the media.

1) TEXTUAL CONTEXT BASED

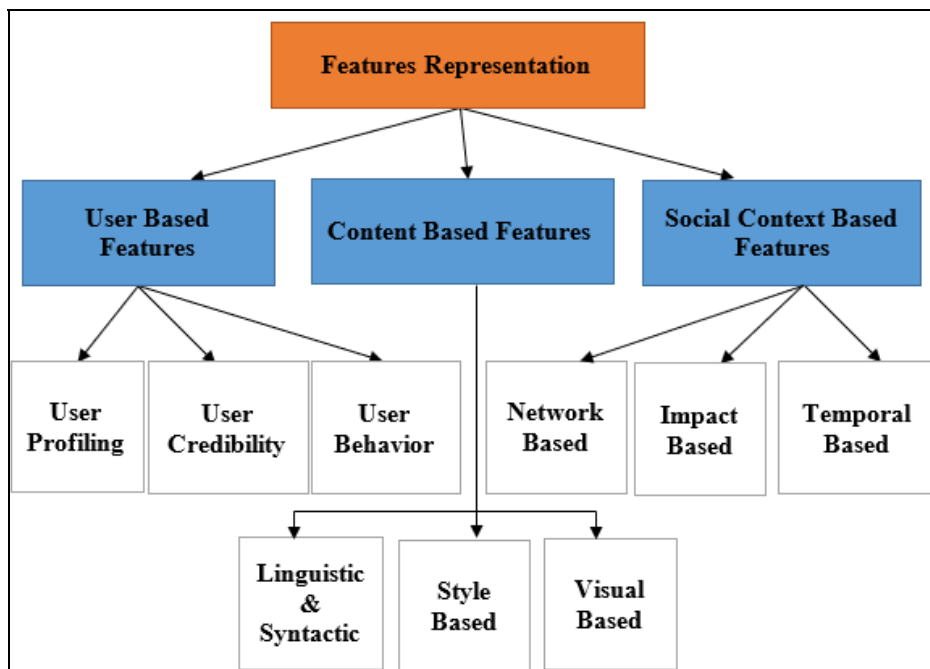Three important methods to make up news content:

FIGURE 3. Feature representation of fake news.

- Source-Where it takes the news or a piece of news getting source, who published it, and source is authentic or not.
- Headline-A detailed summary of the news's quality to entice readers.
- Body Text-It shows the actual story/content of the news. The most common method for detecting false information is to look at the content of the news piece. The substance of a news report is generally separated into two types: textual and visual. Much of the news material is presented in the textual mode, one of these modalities. As previously said, fake news consists of manipulating the audience, and it does so via the use of specific terminology. Non-fake news, however, is usually transferred to a separate language list since it is more legitimate. Attribute-based language characteristics and structure-related language features are two common categories.

2) ATTRIBUTE-BASED LANGUAGE FEATURES
They involve the ten parallel aspects of content style's linguistic elements. These aspects involve volume, uncertainty, objectivity, emotions, diversity, and readability [24]. Although attribute-based language characteristics are generally extremely important, explainable, and predictable, they are often useless in assessing deception style compared to structure-based features. Furthermore, attributed features require extra resources for deception detection, which may take longer and significantly focus on correct feature evaluation and filtering.

3) STRUCTURE-BASED LANGUAGE FEATURES
Content style is defined by structure-based linguistic properties and must have four levels of language: the first one is lexicon, the second is semantics, then discourse and syntax. Structure-related features are also technique-oriented features because most quantification depends on NLP-based methods. The critical challenge at the lexical level is identifying the frequency statistics of a word(s), letter(s), or other entity, which may be done correctly by applying n-gram models. Part-of-Speech (POS)-taggers execute shallow syntax tasks at the syntax level, making tagging and assessment of POS easier. Probabilistic Context-Free Grammars (PCFG) analyses Context-Free Grammars (CFG) by performing deep syntax level operations with parse trees. On the semantic level, word count (WC) and linguistic inquiry are also utilized to create semantic classes for semantic features.

E. PROBLEM FORMULATION
Developing a Spark distributed cluster-based environment for efficiently detecting fake news articles via a supervised learning paradigm necessitated solving two sub-problems. First, our model needed to learn how to recognize and seize necessary information in lengthy and textual news articles for categorizing the association between news item titles and related meta descriptions.

F. RESEARCH OBJECTIVES
In the first section of this research, we examine the effectiveness of Recurrent Neural Networks (RNN) in modeling news articles to identify the link between an article's body content and its title. As part of our research, we use the dataset made available for the FNC-1 competition to train and assess a classifier. We want the classifier to be able to do the following.
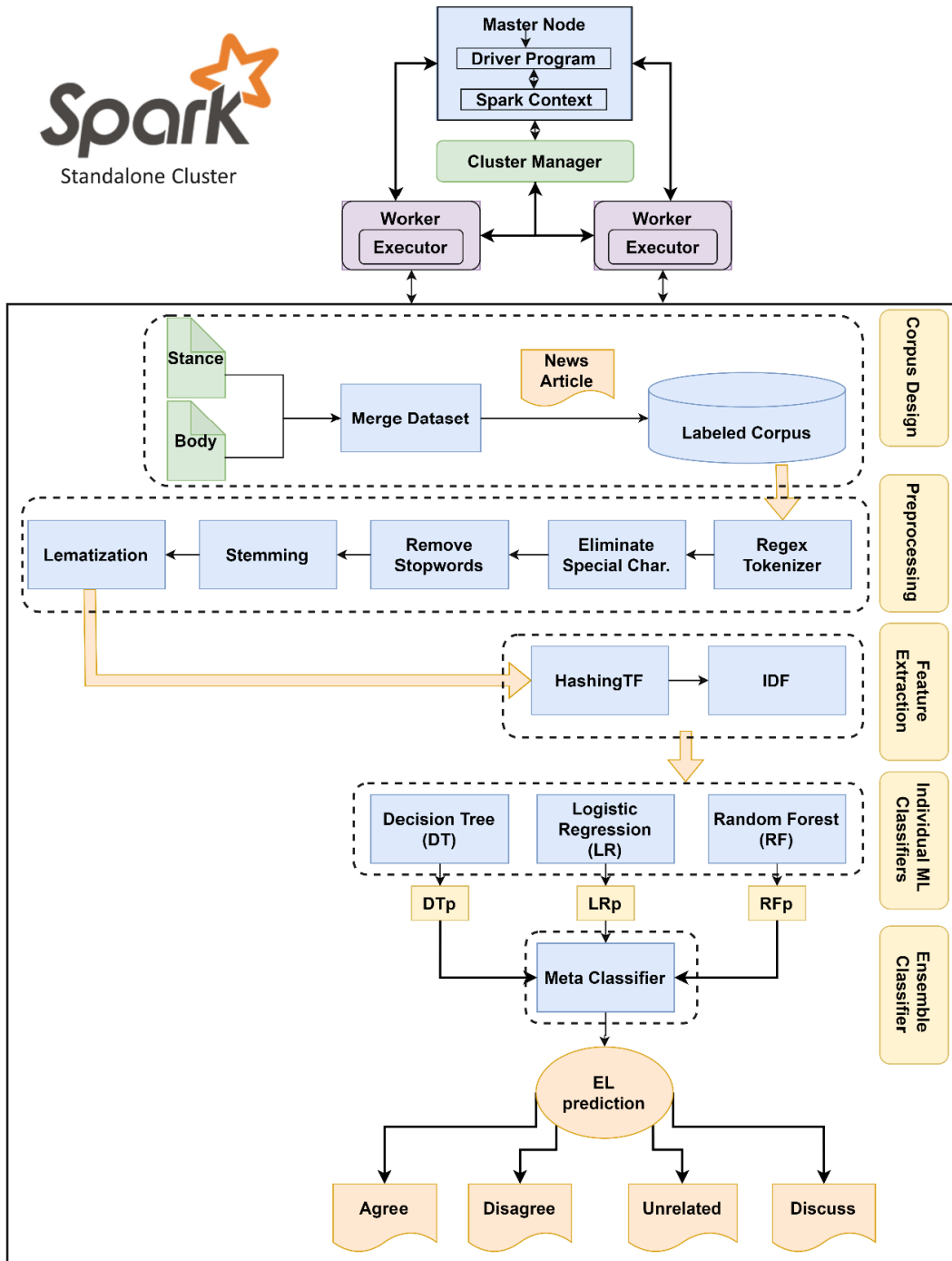
FIGURE 4. Graphical representation of proposed approach.

●Use the Spark framework to research, assess, and compare several machine learning classification techniques on four classes from the FNC-1 dataset.

●Given a title and an article, determine if the article agrees with, disagrees with, discusses, or is irrelevant to the assertion made in the headline.

● To propose an efficient, systematic, and functional approach based on machine learning algorithms for detecting fake news using Spark and to design an efficient stacked ensemble classifier for fake news detection.

In an experiment, we demonstrate that the recommended method can accurately identify fake news and beats current state of the art algorithms.

G. PAPER LAYOUT

The remaining paper contains the following sections. Related work is reviewed in section II. The dataset used for

experimentation and preliminaries is discussed in section III. The experimental results and discussion are articulated in section IV. Finally, section V presents a conclusion, and future work.

II. LITERATURE REVIEW

This section provides an overview of the previous research's difficulties in identifying fake news. To identify fabricated news stories, it is necessary to do rumor detection and identification. It is important to distinguish between Real and fake news since both are based on deliberate fabrication. Fake news identification is particularly difficult when detecting news based on characteristics. Tweets and social context can be used to generate features. As a result, we assess prior work based on single-modality and stance identification.

A. TEXTUAL CONTENT BASED

Most earlier news identification studies relied mainly on textual elements and user metadata. Text based features are statistically extracted from message text content and have been extensively discussed in the literature on fake news identification. The textual component extracts unique writing styles [15], [19], [20] and emotional sensations [18] that are prominent in fake news.

Network connections, style analysis, and individual emotions have all been proven to contribute to detecting fake news [19]. After reading these posts, [20] explored the writing style and its effects on readers' viewpoints and attitudes. Emotion is a significant predictor in many fake news detection studies, and most rely on user positions or simple statistical emotional features to convey emotion. In [15] authors introduced a novel dual emotion-based method for identifying fake news that can learn from publishers' and users' content, user comments, and emotional representation. Reference [25] employed an ML model for identifying fake news that employs convolution filters to distinguish between different granularities of text information. They investigated the issue of posture categorization in an innovative approach to consumer health information inquiries and achieved 84% accuracy using the SVM model.

B. SOCIAL CONTEXT BASED

User generated social media interactions with news stories may give additional information, in addition to aspects directly relevant to the substance of the stories. In [26] authors proposed a novel approach employing a knowledge graph to identify fake news based on actual content. A graph-kernel-based approach used be [27] to discover propagation patterns and attitudes. On the other hand, social context features are difficult to gather because they are loud, unstructured, and time-consuming [28].

C. STANCE DETECTION OVERVIEW

From a broad viewpoint, stance detection can be elaborated as the problem of determining an author's or text's point of view concerning a specified target, such as a single topic, headline, or even a person [15], [29]. Consequently, there are three factors and a machine learning based categorization technique to determine how the comparison occurs. The group's titles (for example: help, against, for, or neutral) are determined by the issue. Political arguments [30], [31], articles [32], [33], and even internal company dialogues [25], [34], which stretches a wide range of fields may be referred to as categories. Detecting the stretch of Tweets or short texts such as hearsays [35] or microblogging accounts has gotten much attention in opinion mining. "Hillary Clinton" as a celebrity, "Atheism" as a specific issue, or the profess that "E-cigarettes are safer than regular cigarettes" are examples of objectives presented in the available datasets. Shared tasks for providing such datasets and promoting research have emerged in several languages.

The sub-task for exposing stance in Tweets [26] was presented at SemEval-2016, with roughly 5,000 tweets in English, including five familiar subjects. The task has initiated a variety of approaches, including conventional techniques (for example, KNN [36], SVM [22], or essential attributes given by methods [34]) and deep learning approaches (e.g., BiLSTM [37], Bidirectional Conditional Encoding [27], [34]). Furthermore, public datasets, for instance, the Multi-Perspective Consumer Health Query dataset [38] dedicated to exposing the stance of sentences taken from high-quality articles on five separate assertions. Like "Sun exposure causes skin cancer," the dataset is available to work on the development of new and exciting work. It contains an in-depth examination of various approaches to the two goals listed above. The need for well-interpreted data in languages other than English has rapidly increased notation efforts and collaborative tasks aimed at furthering research. There are efforts like Stance-Cat, an aim for identifying attitudes in Spanish and Catalan tweets [39], a proposal and database of brief statements in Russian online forums [40], and even projects that integrate several languages [41].

A group of volunteers from industry and academia launched the Fake News Challenge in December 2016 [10]. Using Machine Learning, Natural Language Processing (NLP), and Artificial Intelligence (AI), this competition aimed to encourage the development of technologies that could assist human fact-checkers in detecting deliberate deception in news reporting as a first step, the organizers decided to research what other media outlets have to say about the topic. Consequently, they decided to introduce the event with a stance detection challenge in the first round of competition. The organizers collected data on headlines and body text before the event. In the competition, they asked participants to create classifiers that could reliably classify a body text's viewpoint on a given headline into one of four categories: "disagree", "agree", "discuss" or "unrelated". On this task's test set, the top three teams achieved accuracy rates greater than or equal to 80%. The top team's model combined Gradient Boosted Decision Trees and Deep Convolutional Neural Networks.

D. MISLEADING HEADLINES

Identifying misleading headlines in this research required classifying each article's treatment of the assertion made in the title into one of four categories: (a) agrees, (b) discusses, (c) disagrees, and (d) irrelevant (headline and different topic discussed in body text). As a result of the proliferation of annotated corpora and the increased use of new technologies to combat the fake news pandemic, a new obstacle has recently presented itself to the field of fake news analysis [8]. In this setting, several research challenges and competitions are presented. The most recent and important ones are then dissected in great detail. The evolving dataset [18] was used to create the fake news Challenge6 (FNC-1) [42]. The goal of FNC-1 is to serve as a benchmark for research into AI-based technologies, machine learning, and natural language processing as they apply to the detection of false news. The planners decided to begin with stance disclosure to finish this macro-challenge. The FNC-1 dataset, which included over 75,000 instances labelled as either "agreeing," "discussing," "disagreeing," or "unrelated," was made publicly available. Given the headline "Robert Plant Ripped up $800M Led Zeppelin Reunion Contract," the following excerpts illustrate the categories mentioned, as annotated by the barometer in the FNC-1 dataset.

Body content that conforms to the headline is an instance of agree class. These topics might be discussed in a discussion class: The article's main body addresses the same issue as the title, but does not take a position on the matter. For instance, when comparing the headline and body content, one might say they belong to different classes. The FNC-1 competition had 200 entries, the top 10% of which averaged 82% relative points. The group developed a basic criterion using just hand-coded features and a Gradient Boosting Classifier, both freely accessible on GitHub. Top systems were UCLMR [43], Talos [44], and the Athene system [23]. The CNNs utilised by Talos [44] were one-dimensional, active at the word level, and trained using Google News topic vectors for the article's main body and title. The data from the CNN is then fed into a multi-layer perceptron (MLP) model that generates one of four possible classes of results. Next, it undergoes a comprehensive, start-to-finish training process. The system won the FNC-1 competition with its superior performance using the CNN-MLP combo. In recent trials, several research have employed FNC-1 with encouraging outcomes. For instance, [45] suggested a tree-like structure for the linked classes by combining the existing disagree, agree, and discuss ones. This approach uses a two-layer neural network to learn a hierarchical representation of classes, achieving a weighted accuracy of 88.0%.

Additionally, scholars built a stance detection model using accomplishment transfer learning on a Roberta Deep Bidirectional Transformer Language Model. They achieved a weighted accuracy of 90.01% by employing Bidirectional Cross Attention between claim article pairings via pair encoding with self-attention [46]. Further work should be done on posture identification problems, such as linking a news title and article content, outside the FNC-1 Challenge and dataset. Several writers have compiled claims and criticisms [21], [47] to help with identification. Some analytic effort is devoted to "argument mining," in which the headline presents an argument not supported by the content. While argument mining is effective in solving the problem of posture identification, other tasks that discover semantic relationships within the text, such as inconsistency detection [48], contrast detection [49], and synthesis detection [50], may also be useful. Mishra et al. provided a comprehensive taxonomy for spotting false news, outlining the many forms of disinformation and what sets them apart. Multiple mechanisms exist to track down those who propagate false information. Multiple liar, false news, and corpus datasets have been used to compare traditional machine and deep learning techniques. This study demonstrated that deep learning methods outperformed more conventional machine learning strategies. Bi-LSTM outperforms the competition in detecting bogus news with an F1 score of 96.

In [43] authors introduced the Multi-integrated Domain Adaptive Supervision (MIDAS) system to automatically choose the model that best fits a particular collection of data drawn from random distributions. By using local smoothness as a proxy for accuracy and the relevance of training data, MIDAS can increase generalization accuracy across nine distinct fake news datasets. MIDAS has a larger than 10% success rate in recognizing bogus news linked to COVID-19, compared to other labelling methods [43]. The results of the literature review were summarized in Table 1.

III. PROPOSED METHODOLOGY

This section describes a comprehensive detail about the proposed approach. The proposed approach comprises multiple steps of data analysis, feature extraction, single classifier, and the ensemble classifier classification, as shown in Figures 4. The challenge of fake news in stage 1, a particular purpose and dataset is presented to handle the difficulty of identifying fake news. The challenge's primary motivation is to build a semi-automated pipeline that examines the attitude of several news items on a specific topic. Thus, the dataset comprises occurrences with a title, article body, and one of the four labels "Disagree", "Agree", "Unrelated", and "Discuss". Figure 4 summarizes our proposed approach, which consists of the steps to achieve fake news classification by solving multi-class labels. The first part explains the corpus creation technique by combining stances and bodies based on news article ids. The second phase describes the preprocessing processes done on news article text. The third phase demonstrates techniques to feature selection or dimensionality reduction. The fourth stage describes each ML and ensemble model used in this study. Finally, the last phase outlines this study's various ensemble learning models. We divide the dataset into two parts for experiments: training and testing. The training dataset comprises 75% of the data, whereas the testing dataset contains 25%.

TABLE 1. Literature review summary.

Ref	Dataset	Contribution	Acc %	Challenges
2020 [1]	Twitter-BR, btv-lifestyle, Fake-Br-Corpus, Fake-news-Data1 and Fake-Or-Real news	1. Detecting the fake news in several languages 2. Each dataset was evaluated separately using various feature extraction techniques like; Word-2-Vec, DC-distance technique, etc.	79	To improve accuracy, ELMO embedding is preferred over DCDistance, as DCDistance does not perform noticeably better than other bag of words machine learning techniques.
2019 [2]	Italian Facebook dataset containing 50,000 incorrect information and 300,000 official media news items	A classification model with multiple important feature extraction techniques.	91	The ineffectiveness of the component to be comprehended, which influences the information, showed a lack of comprehension.
2020 [3]	Weibo and Twitter datasets are used	CNN is trained on both labelled and unlabeled samples. Model validation is accomplished using five-fold cross-validation.	90	Only 1,111 Twitter postings and 816 Weibo posts were included in the dataset. That are limited for training deep learning model, Machine Learning model such as LR, perform well in limited dataset
2019 [4]	US election related 2282 news articles	New features for training classifiers have been added.	85	There is only a small amount of data available. The work is extended to implement a better model to make better predictions about fake news by combining many datasets and deep learning algorithms.
2019 [5]	5800 twitter-tweets	By validating tweets with specific training features, BILSTM-CNN can determine whether they are fake news.	86.12	Text-based characteristics are the only ones that can determine whether something is fake. We found that SVM and Random Forest performed significantly better in the textual dataset than the deep learning model.
2020 [6]	PolitiFact dataset, BuzzFeed dataset	The characteristics are extracted using matrix-tensor factorization. In addition, the retrieved features are passed to XGBoost and DNN.	85.86	Implementing XGBoost, a machine learning-based classifier, has surpassed precision, recall, and accuracy by the Deep Neural Network. Using the real-world dataset and these content- and context-based traits, the researchers can classify news stories in real time (Graph-based dataset).
2019 [7]	Liar-dataset	By classifying news using semantic traits, DSSM-LSTM can detect fake news.	99	The extraction of features based on semantic features was not demonstrated in any way. The work that has been suggested to categorize fake news and to increase accuracy is highly desirable.

A. DATASET

Carnegie Mellon University adjunct professor dean Pomerleau, Joostware, and the AI Research Corporation founder Delip Rao hosted a competition called the Fake News Challenge Stage 1 (FNC-1) to investigate the potential of machine learning and natural language processing in the fight against fake news [27]. This issue was the driving force for the competition, which focused on stance detection. This section

TABLE 1. *(Continued.)* Literature review summary.

2020 [8]	Custom Extracted data using Fake-news-Net tool	Deep learning models such as LSTM, RNN, and GRU were used to investigate binary classification.	LSTM=75 RNN=62 GRU=45	Fake news was defined in this dataset regardless of whether the secondary sources reported on primary document information. Various datasets that differentiate between legitimate and fraudulent news should be used for further research.
2019 [9]	PolitiFact, Snopes and Buzzfeed	This geometric deep learning model uses two convolutional layers and two-dimensional features.	92.7	The findings are promising; the proposed study could be expanded by classifying content related to the topic at hand and identifying fake news at an earlier stage.

provides an overview of the competition dataset, the baseline used by the FNC-1 organisers, and the winning strategies used throughout the competition.

It ensued by turning a news story into a headline, then annotated the title and using the story to show where they stood on the assertion they introduced. For this attitude categorization exercise, we have three possible sets of labels: "for," "against," and "observing." The developing dataset [27] is the basis for the FNC-1 competition dataset. To create the FNC-1 dataset, we randomly match headlines and articles from the emerging dataset depending on their attitude toward the linked allegation. In addition, the headlines and articles are separated into related and unrelated groups. Second, and more difficult, the collection of connected headline-article pairings is further split into the three classes disagree, agree, and discuss, allowing for supervision of the job of evaluating the attitude of an article relative to the assertion presented in the associated headline. There are 49,972 headline-article pairs in the training set of the FNC-1 dataset, and another set of pairs in the test set. There are 1,689 distinct headlines and 1,648 unique articles used to build the headline-article pairings that make up the training set. The test set includes 904 distinct articles and 894 unique headlines. Seventy-three percent are classified as unrelated, 7.4 percent as agreeing, 1.7 percent as disagreeing, and 17.8 percent as debating. About 72.2 percent of the test data is irrelevant; 7.4 percent is in agreement; 2.7 percent is in disagreement; and 17.6 percent is up for discussion. The training set has 40,350 headline-article sets, the hold-out set has 9,622, and the claim set has 25,413 sets.

B. CORPUS DESIGN

The dataset FCN-1 has four distinct classes (agree, disagree, discuss, unrelated). In pre-processing, labels are encoded into numeric target values and perform some pre-processing steps. Preprocessed data is split into 75% data for training and 25% for testing.

This study used the FNC-1 dataset, consisting of two CSV files, including stances and body corpora of text news stories written in English. Collecting news stories from multiple sources is difficult due to a lack of linguistic resources. Furthermore, annotating these news pieces based on their contents necessitates specialist expertise, a significant amount of time, and substantial money. As a result, augmented corpus design is the only way to conduct fake news detection research. Our augmented corpus is created by combining 49972 stances with 1683 bodies based on ids. The corpus has four distinct classes (agree, disagree, discuss, unrelated). It contains 8909 discuss stances, 36545 unrelated stances, 3678 stances, and 840 disagree stances. After gathering headlines and articles in one column, the final corpus contains text and stances.

C. PRE-PROCESSING

Data mining relies heavily on pre-processing. It converts inconsistent and incomplete raw data into a machine-readable representation. Various text preprocessing activities were conducted on the FNC-1 dataset. To complete these tasks, NLP approaches such as character conversion to lowercase letters, stop word elimination, stemming, and tokenization, as well as algorithms from keras library were used. Stop words, which comprise words like "the, of, there," etc., are the most commonly used words in our daily language and typically have relatively limited significance in terms of the entire context of the phrase. By removing the stop words, we save time and space that would otherwise be consumed by the useless phrases mentioned before. Words with comparable meanings may appear in the text many times. For example, "eating" in any sentence will become "eats". Reducing the language to its most basic form can help if that's the case. This operation, known as stemming [51], uses an open-source version of the NLTK's Porter stemmer method. Few preprocessing steps are as follows:

1) **Stop Word Removal:** Languages commonly use a group of terms collectively known as "stop words." The words "a," "the," "is," and "are" are all examples of stop words in English. Stop words are common in text mining and natural language processing (NLP) to weed out overused words and thus contain little useful information. NLTK provides the stop word dictionary

in this instance. To begin, the text is cleaned up by removing all stop words. It is possible to remove stop words from the text because they are more common and carry less useful information. Some common stop words include the conjunctions 'and' 'or' and 'but'. Pre-processing data is essential in natural language processing because processing these less frequently used full words consumes a significant amount of time.

2) **Punctuation Removal:** The grammatical context of a sentence is provided by natural language punctuation. A comma, for example, may not add anything to the understanding of the statement.

3) **Link Removal:** This step removes hypertext links from social media posts. Regular expressions are used to do this.

4) **Lemmatization or stemming:** Either lemmatization or stemming is done during this step. The NLTK's WordNet Lemmatizer is used for lemmatization, while the NLTK's Snowball Stemmer implementation is used for stemming, based on the Porter2 stemming algorithm [52].

5) Apart from the above-mentioned pre-processing stages, every social media post must go through. Reply removal: Words beginning with @ (primarily used for Twitter replies) are eliminated in this phase. Regular expressions are also used to do this.

6) **Lowercase transformation:** Every word is converted to lowercase in this phase to account for variances in capitalization.

D. FEATURE EXTRACTION

Feature extraction transforms raw data into numerical features that can be further processed while preserving the original data set's information. It is more effective than just using raw data to train a machine.

1) HASHINGTF

The mapped indices are then used to calculate the phrase frequency. Bypassing the need for a term-to-index map, which can be time consuming and expensive for large corpora, this method is less susceptible to hash collisions [45], where multiple raw features are hashed into the same term. HashingTF maps a series of phrases to their word frequencies using the hashing method. Using Austin Appleby's Murmur Hash 3 algorithm, we can now compute the term object's hash code value (MurmurHash3 $\times$ 86 32). Since the hash function is translated to a column index using a simple modulo, the features would not be evenly mapped to the columns if the numb-Features input was less than a power of two. The HashingTF transforms a set of terms into feature vectors of fixed length. Regarding text processing, a "term set" could be a collection of words. HashingTF employs the hashing technique. A hash function transforms a raw attribute into an index (term). Murmur-Hash-3 is the hash function in use here. The mapped indices are then used to calculate the phrase frequency. When working with large datasets,

avoiding creating a global term-to-index map is preferable because doing so can be time-consuming and expensive. However, this method is vulnerable to hash collisions, which occur when different raw features are hashed into the same term. Increasing the number of buckets in the hash table to reduce the likelihood of collisions is recommended. A simple modulo determines the vector index on the hashed value, so the feature size should be a power of two. If the feature size is smaller than this, the vector indices will not be evenly distributed. There is a binary toggle parameter that controls the frequency of terms. When this value is true, all nonzero frequency counts are reset to 1. As a result, discrete probability models are built that do not use integer counts but rather binary ones.

2) IDF

Inverse Document Frequency (IDF) is a calculation frequently employed in association with term frequency. The issue with term frequency is that frequent terms are not necessarily the most significant. For example, "content" will appear on every web page. IDF is a method for lowering the weight of frequently occurring words in a corpus (collection of documents). IDF is determined by dividing the total number of documents by the number of documents containing the phrase in the collection. IDF is an Estimator that generates an IDF Model after being fitted to a dataset. Feature vectors (typically created by Hashing-TF or count-vectorizer) are used to scale each IDF model feature [46]. It appears to downplay qualities that are common in a corpus.

E. CLASSIFICATION MODELS AND PARAMETERS SETTINGS

We use the following machine learning techniques to detect irregularities and breakdown of unusual events and investigate the effectiveness of our advanced method:

Random Forest (RF): a supervised learning technique that may be used for classification, retrieval, and other tasks. It generates a few trees to aid in decision-making. It takes a random sample of data, constructs many decision trees to forecast each tree, and then votes on the best option. n-estimators = 200, bootstrap = True, criterion = Gini, min-samples-split = 2, random-state = 0, and min-samples-leaf = 1 are the parameters for our RF method.

Logistic Regression (LR): It is a segregated targeted learning model. A very straightforward ML algorithm differentiates problems such as noise detection, diabetes prediction, cancer detection, etc. LR is used to predict probability of target variability [47]. In our application the parameters of the LR algorithm are Penalty = l2, C = 1.0, reduce rating = 1, solver = lbfgs, max iter = 100 and verbose = 0.

Decision Tree (DT): are extensively used in decision analysis and machine learning [21]. It's a decision-making tool that uses a tree-like graph of decisions and consequences, such as random event outcomes, resource costs, and utility, to make judgments. Internal nodes in a DT express a

condition about an attribute. Each internal node divides into branches depending on the condition's outcome until it reaches a point where it no longer splits and leads to leaf nodes, which indicate the class label that will be applied [48].

Ensemble Classifier: In addition to the custom classifiers, an ensemble technique was developed, which combined the three custom classifiers. The objective is to develop a voting classifier that calculates the weights to apply to each classifier's prediction [53]. The probabilities computed by the classifiers are first stored in a matrix for each training instance, resulting in each training case being linked with a probability vector. The weights are calculated, and the final label is created using this matrix of vectors, which is then fed into a Meta classifier model (0, 1, 2, or 3).

In contrast to the ensemble model, a voting classifier was also constructed to perform simple majority voting among the models' predictions. Ensemble categorization is generally divided into two stages: base-level and ensemble-level. This base predictors employ the HashingTF with IDF received from news articles as input. The output predictions from these base-predictors are fed into ensemble-level models. The ensemble model's main purpose is to improve the overall prediction F1 score by overcoming the shortcomings of the primary predictors. We have used stacking ensemble models for ensemble classification [54].

F. EVALUATION METRICS

The main concern is determining the model's ability to discern true and false news. We used metrics to properly examine the model's efficiency for this difficult challenge. Model selection and implementation are essential but should not take precedence over the rest of the project. Various assessment measures are used to test data to assess the model's capacity to detect false news. Multiple evaluation metrics, such as classification reports (accuracy, precision, recall, F1-score) and confusion measures, may be used to assess machine learning models. The sections that follow go through each of the assessment measures in detail. Preprocessing and other ways of gathering fake news data are loaded into a strong algorithm, producing incredible results [49].

Observations that match the predictions made by the model are true positives and negatives, respectively, and are marked in green. Because we would want to cut down on both types of errors, the ones we are trying to minimize are marked below. These phrases don't make a lot of sense. So, we can check our understanding by dissecting each statement.

A True Positive (TP) is a correctly anticipated positive result when the actual and projected class values are yes. For instance, if the expected and actual class values suggest that the passenger made it, we know they did. When both the actual and anticipated class values are negative, we say the value is a True Negative (TN). For instance, this passenger did not survive if both the actual and predicted classes suggest

that they did not. When the actual class size differs from the projected class size, false positives and negatives occur.

When the expected class is present, but the real class is not, this is called a false positive (FP). The actual class will be utilized if, for example, it shows that the passenger did not survive but the fore-cast class predicts that the passenger would. In cases when the true class is yes but the predicted class is no, a false negative has occurred. If, for example, the actual class value reveals that the passenger lived whereas the expected class value predicted that they would die, the actual class value would be utilized.

To verify the usefulness of the model, the following assessment criteria are used:

Precision is the proportion of actual test results that were predicted correctly. This is calculated by dividing the number of correct predictions by the number of incorrect ones.

$$Acc = \frac{TP + TN}{TP + FN + TN + FN}$$

Precision: To calculate a classifier's precision, divide the number of positive outcomes by the number of positive predictions.

$$Pr = \frac{TP}{TP + FP}$$

Recall: The total number of positive outcomes divided by the total number of predicted positive outcomes is used to determine recall.

$$Re = \frac{TP}{TP + FP}$$

F1-score: It is a great way to test accuracy and recall simultaneously. This value is used to gauge accuracy and recall.

$$F1 = \frac{2 \times (Precision \times Recall)}{Precision + Recall}$$

The accuracy of a prediction may be measured with the use of a classification report (CR). Correct and incorrect classifications for each category are utilized to determine the totals. False positive (FP), true negative (FN), and false positive and negative (FP/FN) are widely used in the classification report's construction (FP&N). Several metrics may be used to evaluate a model's efficacy, but accuracy is often prioritized. For example, it incorporates a wide range of assessment tools including as (accuracy, precision, recall, F1 score, and support.) The backing indicates the number of occurrences for each class [50]. It represents how much information out of the total possible may be calculated with high precision. Number of courses where just the best features were recalled. An equation may be used to depict this. To get the F1-score, we add the percentage of correct predictions and the number of correct recalls. The table summarizes the mean weighted recall and accuracy for a certain sample. The F1-score for this model is 1, which means it is ideal. "Support" refers to the number of class occurrences in a given dataset. The word "accuracy" refers to the proportion of correct predictions relative to the number of potential ones.

TABLE 2. Proposed approach results.

Proposed Approaches	Precision %	Recall %	F1-score %	Accuracy %
HashingTF-IDF_Ensembler	92.03	92.45	92.25	93.45
Lr_Bigram	87.02	88.01	87.06	88.45
Lr_Glove	63.12	73.25	62.45	73.25
Lr_Cv+IDF	83.12	84.25	83.26	84.54
Lr_Uni+Bi+Tri +Cv+IDF+Chiseq	82.12	83.35	82.45	83.45
LR_CV	88.25	88.45	87.35	88.45
SVM_CV	91.25	91.24	90.45	91.75
LR_HashingTF-IDF	93.14	93.45	93.45	93.78
Svm_HashingTF-IDF	90.45	90.75	90.32	90.75
Lr_Trigram	82.01	83.45	82.64	83.47
Lr_Uni+Bi+Tri	88.12	88.36	87.64	88.64
Lr_Uni+Bi+Tri+16000	82.12	83.47	82.65	83.78

IV. EXPERIMENTS AND RESULTS

A. CLASSIFICATION RESULTS

The experimental results of Term Frequency-Inverse Document Frequency (TF-IDF) and HashingTF feature extraction techniques with ensemble models are presented in Table 2. The results using HashTF and IDF features regarding accuracy, precision, recall, and F1-score are 93.45%, 92.03%, 92.45%, and 92.25%. The results from LR_HashingTF-IDF is 93.45%, and it's a highest as compared to all other experimental. Furthermore, Bigram Logistic Regression exhibits 88.45% accuracies, 87.02% precision, 88.01% recall, and 87.06% F1-score. We also performed experiments using glove word embedding. We used the glove embedding technique with logistic regression. However, the glove with logistic regression model results is not so high but quite well with accuracy scores of 73.25% and 63.12%, 73.25%, 62.45% as the precision, recall, and F1-score. To make a broader comparison, we include features of the count vectorizer technique. The features of the count vectorizer were passed to logistic regression to detect fake news. Using the count vectorizer technique, the logistic model achieved 88.45% accuracy, 82.12% precision, 88.45% recall, and 87.35% F1 score. Moreover, we merged the count vectorizer and TF-IDF features to obtain better results, but we failed to avail improved results due to the high computational cost. The correctness, precision, recall, and F1-score using count vectorizer and TF-IDF features with logistic regression are 84.54%, 83.12%, 84.25%, and 83.26%. We also employed the Support Vector Machine (SVM) model to testify its abilities using count vectorizer features, and the SVM model gets improved results with 91.75% accuracy, 91.25%

precision, 91.24% recall, and 90.45% F1-score. As compared to LR with count vectorizer, the SVM obtained high results. We also employed LR and SVM models with HashingTF-IDF features. The results of LR with HashingTF-IDF are better than the SVM model. Compared to LR, the SVM model with HashingTF-IDF achieved 90.75% accuracy.

The LR model with HashingTF-IDF obtained 93.78% accuracy, which is higher than the SVM model's accuracy. At the end we utilized Trigram, Unigram + Bigram + Trigram, Unigram + Bigram + Trigram + 16000 limited top features and Unigram + Bigram + Trigram + Cv + IDF + Chiseq feature with Logistic Regression to efficiently detect fake news. The LR with Trigram obtains significant results: accuracy is 83.47%, precision is 82.01%, recall is 83.45%, and F1-score is 82.64%. While compared to individual Trigram features, the LR model with Uni, Bi, and Trigram obtained better results with 88.64% accuracy. However, when running tests with Uni, Bi, Trigram, and 16000 limited top features, the LR model obtained less accuracy, which is 83.78%. Ultimately, we tried to merge all the features Unigram + Bigram + Trigram + Cv + IDF + Chiseq, applied LR on these features, and obtained promising results with 83.45% accuracy and accuracy and 82.45% F1-score.

The Figure 5 (a) shows the classification report of ensemble model. The support presents the number of instances of each class in testing set. 12,403 instances are used for testing data. We used weighted accuracy to calculate the precision, recall and F1-score because it deals with the class imbalance problem. The mean average precision, recall, F1-score of all classes is calculated using macro average, while weighted

TABLE 3. Comparative analysis of proposed and baseline approaches.

System	F1 Score %				F1-score%	Acc%
	Agree	Disagree	Discuss	Unrelated		
Existing Approaches						
[23]	0.487	0.151	0.780	0.996	0.604	0.894
[24]	0.588	0.667	0.765	0.997	0.754	-
[42]	0.479	0.114	0.747	0.989	0.583	0.884
[44]	0.539	0.354	0.760	0.994	0.582	0.890
[45]	0.674	**0.813**	0.839	0.997	**0.831**	0.937
[55]	**0.737**	0.552	0.855	0.991	0.784	0.937
[56]	0.703	0.534	**0.853**	**0.994**	0.771	**0.936**
[57]	0.723	0.625	0.873	0.993	0.803	0.943
Proposed Approaches						
LR_CV_features	0.634	0.445	0.766	0.934	0.877	0.880
SVM_CV_feature	0.712	0.636	0.814	0.956	0.905	0.911
LR_TF-IDF_features	b	0.785	0.862	0.962	0.923	0.932
SVM_TF-IDF_features	0.711	0.645	0.801	0.952	0.902	0.903
Lr_bigrams	0.614	0.414	0.746	0.934	0.875	0.884
Lr_Tri_gram	0.533	0.325	0.647	0.903	0.826	0.833
Lr_Uni+bi+tri	0.645	0.436	0.777	0.938	0.875	0.886
Ensemble model	0.793	0.743	0.853	0.968	**0.924**	0.934
Lr_glove	0.000	0.000	0.002	0.843	0.623	0.735
Lr_cv_IDF	0.553	0.403	0.673	0.912	0.834	0.842
Lr_uni+bi+tri+cv+IDF+chisq	0.552	0.424	0.652	0.902	0.829	0.832
Lr_Uni+bi+tri+16000f	0.552	0.423	0.662	0.902	0.821	0.831

average is the total number of TP divided by the entire number of objects in all classes. Macro average stands for mean average. The weighted average score is higher due to the class unbalancing in the dataset. We also construct the ensemble model's confusion matrix, as shown in Figure 5(b). A confusion matrix, also known as an error matrix, is a table that visually depicts the performance of a supervised classification machine learning system. Figure 5 (b) shows that the model made multiple incorrect classifications. The ensemble model's ultimate accuracy on testing data is 93%.

B. PERFORMANCE COMPARISON OF DIFFERENT APPROACHES

The comparative analysis of proposed approaches with various baseline approaches is presented in Table 3. The bold values manifest the highest achieved score of proposed and baseline approaches. The experimental setting of proposed approaches resembled the baseline. It is shown in the table 3 that the proposed approach with TF-IDF features and LR

model outperforms the baseline highest F1 score, which is 83.10%, while the proposed approach obtained the highest F1 score of 93.84%. In addition, dealing with the class-wise score, the baseline approach of [46] exhibits the best score for Agree class with 73.76%. The proposed approach with TF-IDF features and LR model achieved the highest agree with class score of 80.23%. The proposed approach outperforms the baseline regarding the F1 score, with the highest F1 score of 92.45% and improved 9.35 %.

C. DISCUSSION

The FNC-1 dataset, which contains 49,972 headline articles and four distinct categories, was used to achieve the investigation's objectives, and obtain the desired results (discuss, agree, unrelated, and disagree). The proposed system comprises numerous components, such as data pre-processing, visualization, exploratory analysis, feature extraction, and classification using machine learning strategies. We proposed classifying data with an ensemble model influenced by

```
              precision   recall  f1-score   support

       0.0       0.94      0.97      0.96      9157
       1.0       0.89      0.82      0.85      2147
       2.0       0.82      0.76      0.79       894
       3.0       0.79      0.69      0.74       205

  accuracy                          0.93     12403
 macro avg       0.86      0.81      0.84     12403
weighted avg     0.92      0.93      0.92     12403
```
(a)
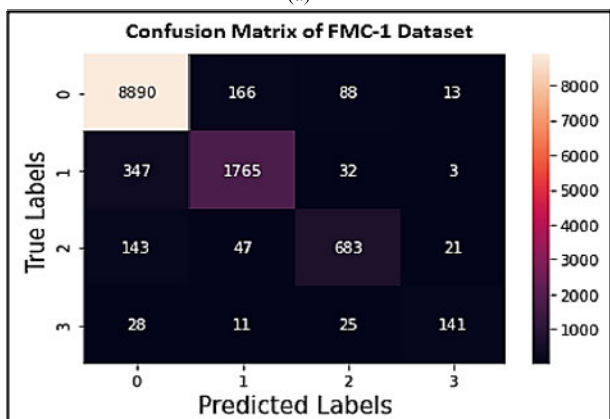
(b)

FIGURE 5. (a): Classification report of final ensemble model (b): Confusion matrix of final ensemble model.

machine learning in real time during the experiment. As a direct result, a more rapid interpretation of the findings is now possible. Instead of just one, two, or three different classification methods, the proposed ensemble model employs three distinct machine learning approaches (Random Forest, Logistic Regression, and Decision Tree). This ensemble model was created as part of our efforts to improve our previous investigations into identifying and categorizing fake news.

Several different factors are influencing the current situation. Several experiments are being carried out using the Apache Spark framework to handle big data and perform classification task. These experiments were carried out to improve our ability to detect fake news. As a result of these experiments, our ability to recognize hoaxes and other forms of disinformation should be enhanced. The model's was one of the aspects considered during the evaluation process for this particular piece of research. The model's accuracy was also considered part of the evaluation process, in addition to its performance compared to five other distinct criteria. Different evaluation metrics include accuracy, precision, recall, the F1-score, and the confusion matrix to test the model's performance.

PySpark was chosen because it uses RDD, significantly accelerating computation processing. As a result, the computations were finished significantly faster than they otherwise would have been. This was the essential consideration in

deciding whether or not to employ PySpark. Compared to the other approaches utilized during this inquiry and the previous baseline studies, the suggested ensemble model had the greatest F1 score. The proposed ensemble model exhibits the highest F1 score compared to the existing baseline studies and other approaches used in this study. This model achieved the highest F1 score of 92.45% due to the features of HashingTF-IDF that were added during development. We boosted our F1 score by 9.35%, which is a sufficient gain to prove the novelty of this research.

In the future, one of our long-term goals is to use Spark to implement deep learning models in a multi-agent distributed learning environment. These algorithms will be used to detect instances of fake news. As a result, we can assess the effectiveness of a wide range of machine learning and deep learning algorithms on a diverse set of fabricated news stories. Furthermore, we intend to create a featured ensemble of different embedding techniques alongside different machine learning and deep learning models capable of accurately recognizing and categorizing various hoaxes and fake news. This will be done so that we may better understand how to spot false news, which will not only aid in understanding the patterns of detecting hoax or fake news but also in developing a cutting-edge real-time fake news detection system.

V. CONCLUSION

Headline stance checker has been indicated to be a helpful method for exposing falsehood in the news, particularly when a headline is contrasted to its content body. To demonstrate the applicability of the headline stance checker, various tests were conducted in the context of an existing assignment (Fake News Challenge FNC-1). The stance of a headline had to be categorized into one of the following classes: disagree, agree, unrelated, and discuss. The studies included verifying each of the suggested classification steps separately and the overall method is evaluated by comparing the state-of-the-art in this job. In this study, researchers used the dataset FNC-1, which has categorized fake news into four categories, while using big data technology (Spark) to perform machine learning analysis for assessment and comparison with other state-of-the-art approaches in fake news identification. The suggested approach created a stacked ensemble model and experimented with it on a distributed Spark cluster. We used N-grams, HashingTF-IDF, and count vectorizer for feature extraction, followed by the suggested stacked ensemble classification model. Compared to the baseline techniques' results, the suggested model has a high classification performance of 92.45% in F1-score. The suggested model outperforms the previous baseline techniques and improves the F1 score significantly. The suggested ensemble model improves the F1 score by 9.35%.

A. RECOMMENDATIONS FOR FURTHER WORK

We currently work with a supervised approach, but researchers can work with unsupervised fake news detection in the future. This proposed work can also be extended

using various neural network-based models, which are more sufficient for unsupervised fake news detection. Spark takes too much training time due to the standalone cluster. Due to the solitary cluster, Spark takes twice as long to train. In future, researchers can perform experiments on creating a cluster on a different machine. This research may be further stretched by employing various neural network-based models better suitable for unsupervised fake news identification. We will try to build a cluster on a separate computer.

ACKNOWLEDGMENT

REFERENCES

[1] P. H. A. Faustini and T. F. Covões, "Fake news detection in multiple platforms and languages," *Expert Syst. Appl.*, vol. 158, Nov. 2020, Art. no. 113503.

[2] M. D. Vicario, W. Quattrociocchi, A. Scala, and F. Zollo, "Polarization and fake news: Early warning of potential misinformation targets," *ACM Trans. Web*, vol. 13, no. 2, pp. 1–22, May 2019.

[3] Y. Liu and Y.-F.-B. Wu, "FNED: A deep network for fake news early detection on social media," *ACM Trans. Inf. Syst.*, vol. 38, no. 3, pp. 1–33, Jul. 2020.

[4] J. C. S. Reis, A. Correia, F. Murai, A. Veloso, and F. Benevenuto, "Supervised learning for fake news detection," *IEEE Intell. Syst.*, vol. 34, no. 2, pp. 76–81, Mar. 2019.

[5] M. Z. Asghar, A. Habib, A. Habib, A. Khan, R. Ali, and A. Khattak, "Exploring deep neural networks for rumor detection," *J. Ambient Intell. Hum. Comput.*, vol. 12, no. 4, pp. 4315–4333, Apr. 2021.

[6] R. K. Kaliyar, A. Goswami, and P. Narang, "DeepFakE: Improving fake news detection using tensor decomposition-based deep neural network," *J. Supercomput.*, vol. 77, no. 2, pp. 1015–1037, Feb. 2021.

[7] S. S. Jadhav and S. D. Thepade, "Fake news identification and classification using DSSM and improved recurrent neural network classifier," *Appl. Artif. Intell.*, vol. 33, no. 12, pp. 1058–1068, Oct. 2019.

[8] A. Vereshchaka, S. Cosimini, and W. Dong, "Analyzing and distinguishing fake and real news to mitigate the problem of disinformation," *Comput. Math. Org. Theory*, vol. 26, no. 3, pp. 350–364, Sep. 2020.

[9] F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, "Fake news detection on social media using geometric deep learning," 2019, *arXiv:1902.06673*.

[10] M. H. Goldani, S. Momtazi, and R. Safabakhsh, "Detecting fake news with capsule neural networks," 2020, *arXiv:2002.01030*.

[11] S. Shellenbarger, "Most students don't know when news is fake, Stanford study finds," *Wall Street J.*, vol. 21, 2016.

[12] D. Pierson, "Facebook and Google pledged to stop fake news. So why did they promote Las Vegas-shooting hoaxes?" Los Angeles Times, Oct. 2017.

[13] G. Zarrella and A. Marsh, "MITRE at SemEval-2016 task 6: Transfer learning for stance detection," 2016, *arXiv:1606.03784*.

[14] S. Ghosh, P. Singhania, S. Singh, K. Rudra, and S. Ghosh, "Stance detection in web and social media: A comparative study," in *Proc. Int. Conf. Cross-Lang. Eval. Forum Eur. Lang.* Cham, Switzerland: Springer, 2019, pp. 75–87.

[15] A. I. Al-Ghadir, A. M. Azmi, and A. Hussain, "A novel approach to stance detection in social media tweets by fusing ranked lists and sentiments," *Inf. Fusion*, vol. 67, pp. 29–40, Mar. 2021.

[16] S. Somasundaran and J. Wiebe, "Recognizing stances in ideological on-line debates," in *Proc. NAACL HLT Workshop Comput. Approaches Anal. Gener. Emotion Text*, 2010, pp. 116–124.

[17] A. Konjengbam, S. Ghosh, N. Kumar, and M. Singh, "Debate stance classification using word embeddings," in *Proc. Int. Conf. Big Data Anal. Knowl. Discovery*. Cham, Switzerland: Springer, 2018, pp. 382–395.

[18] A. Faulkner, "Automated classification of stance in student essays: An approach using stance target information and the Wikipedia link-based measure," in *Proc. 27th Int. Flairs Conf.*, May 2014.

[19] J. Weedon, W. Nuland, and A. Stamos, *Information Operations and Facebook*. Menlo Park, CA, USA: Facebook, 2017.

[20] A. Vlachos and S. Riedel, "Identification and verification of simple claims about statistical properties," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 2596–2601.

[21] S. N. Shorabeh, N. N. Samany, F. Minaei, H. K. Firozjaei, M. Homaee, and A. D. Boloorani, "A decision model based on decision tree and particle swarm optimization algorithms to identify optimal locations for solar power plants construction in Iran," *Renew. Energy*, vol. 187, pp. 56–67, 2022.

[22] E. Zotova, R. Agerri, and G. Rigau, "Semi-automatic generation of multilingual datasets for stance detection in Twitter," *Expert Syst. Appl.*, vol. 170, May 2021, Art. no. 114547.

[23] S. Mishra, P. Shukla, and R. Agarwal, "Analyzing machine learning enabled fake news detection techniques for diversified datasets," *Wireless Commun. Mobile Comput.*, vol. 2022, pp. 1–18, Mar. 2022.

[24] A. Spark, "Apache spark," *Retrieved January*, vol. 17, no. 1, p. 2018, 2018.

[25] A. Sen, M. Sinha, S. Mannarswamy, and S. Roy, "Stance classification of multi-perspective consumer health information," in *Proc. ACM India Joint Int. Conf. Data Sci. Manage. Data*, Jan. 2018, pp. 273–281.

[26] S. V. Vychegzhanin and E. V. Kotelnikov, "Stance detection based on ensembles of classifiers," *Program. Comput. Softw.*, vol. 45, no. 5, pp. 228–240, Sep. 2019.

[27] C. Silverman, "Lies, damn lies and viral content," Tow Center Digit. Journalism, Columbia Univ., New York, NY, USA, 2015.

[28] S. Harabagiu, A. Hickl, and F. Lacatusu, "Negation, contrast and contradiction in text processing," in *Proc. AAAI*, vol. 6, 2006, pp. 755–762.

[29] S. Mohammad, S. Kiritchenko, P. Sobhani, X. Zhu, and C. Cherry, "SemEval-2016 task 6: Detecting stance in tweets," in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*. San Diego, CA, USA: Association for Computational Linguistics, 2016, pp. 31–41.

[30] B. G. Patra, D. Das, and S. Bandyopadhyay, "JU_NLP at SemEval-2016 task 6: Detecting stance in tweets using support vector machines," in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*, 2016, pp. 440–444.

[31] H. Elfardy and M. Diab, "CU-GWU perspective at SemEval-2016 task 6: Ideological stance detection in informal text," in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*, 2016, pp. 434–439.

[32] I. Augenstein, T. Rocktäschel, A. Vlachos, and K. Bontcheva, "Stance detection with bidirectional conditional encoding," 2016, *arXiv:1606.05464*.

[33] P. Wei, W. Mao, and D. Zeng, "A target-guided neural memory model for stance detection in Twitter," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8.

[34] S. Zhou, J. Lin, L. Tan, and X. Liu, "Condensed convolution neural network by attention over self-attention for stance detection in Twitter," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–8.

[35] M. Taulé, M. A. Martí, F. M. Rangel, P. Rosso, C. Bosco, and V. Patti, "Overview of the task on stance and gender detection in tweets on Catalan independence at IberEval 2017," in *Proc. 2nd Workshop Eval. Hum. Lang. Technol. Iberian Lang. (CEUR-WS)*, vol. 1881, 2017, pp. 157–177.

[36] M. Lai, A. T. Cignarella, D. I. Hernández Farías, C. Bosco, V. Patti, and P. Rosso, "Multilingual stance detection in social media political debates," *Comput. Speech Lang.*, vol. 63, Sep. 2020, Art. no. 101075.

[37] S. Sommariva, C. Vamos, A. Mantzarlis, L. U.-L. Dào, and D. M. Tyson, "Spreading the (fake) news: Exploring health messages on social media and the implications for health professionals using a case study," *Amer. J. Health Educ.*, vol. 49, no. 4, pp. 246–255, Jul. 2018.

[38] B. Riedel, I. Augenstein, G. P. Spithourakis, and S. Riedel, "A simple but tough-to-beat baseline for the Fake News Challenge stance detection task," pp. 1–6, May 2018, *arXiv:1707.03264*.

[39] Q. Zhang, S. Liang, A. Lipani, Z. Ren, and E. Yilmaz, "From stances' imbalance to their hierarchical representation and detection," in *Proc. World Wide Web Conf.*, May 2019, pp. 2323–2332.

[40] C. Dulhanty, J. L. Deglint, I. B. Daya, and A. Wong, "Taking a stance on fake news: Towards automatic disinformation assessment via deep bidirectional transformer language models for stance detection," 2019, *arXiv:1911.11951*.

[41] B. Pouliquen, R. Steinberger, and C. Best, "Automatic detection of quotations in multilingual news," in *Proc. Recent Adv. Natural Lang. Process.*, 2007, pp. 487–492.

[42] M.-C. De Marneffe, A. N. Rafferty, and C. D. Manning, "Finding contradictions in text," in *Proc. Assoc. Comput. Linguistics*, 2008, pp. 1039–1047.

[43] A. Suprem and C. Pu, "MiDAS: Multi-integrated domain adaptive supervision for fake news detection," 2022, *arXiv:2205.09817*.

[44] O. Levy, T. Zesch, I. Dagan, and I. Gurevych, "Recognizing partial textual entailment," in *Proc. 51st Annu. Meeting Assoc. Comput. Linguistics*, vol. 2, 2013, pp. 451–455.

[45] C. Cai and D. Lin, "Find another me across the world - large-scale semantic trajectory analysis using spark," 2022, *arXiv:2204.00878*.

[46] M. D. N. Darji, S. M. Parikh, and H. R. Patel, "Sentiment analysis of unstructured data using spark for predicting stock market price movement," in *Inventive Computation and Information Technologies*. Singapore: Springer, 2022, pp. 521–530.

[47] A. A. H. Ahmadini, "A novel technique for parameter estimation in intuitionistic fuzzy logistic regression model," *Ain Shams Eng. J.*, vol. 13, no. 1, Jan. 2022, Art. no. 101518.

[48] S. Choe, A. Ha, J. W. Jeoung, K. H. Park, and Y. K. Kim, "Macular sector-wise decision tree model for the prediction of parafoveal scotoma not detected by 24–2 visual field test," *Clin. Exp. Ophthalmol.*, vol. 50, no. 5, pp. 510–521, Jul. 2022.

[49] S. Visa, B. Ramsay, A. L. Ralescu, and E. Van Der Knaap, "Confusion matrix-based feature selection," *MAICS* vol. 710, no. 1, pp. 120–127, 2011.

[50] A. Basarkar, "Document classification using machine learning," San José State Univ., San Jose, CA, USA, 2017.

[51] E. Loper and S. Bird, "NLTK: The natural language toolkit," 2002, *arXiv:cs/0205028*.

[52] C. Li, A. Porco, and D. Goldwasser, "Structured representation learning for online debate stance prediction," in *Proc. 27th Int. Conf. Comput. Linguistics*, 2018, pp. 3728–3739.

[53] R. Sandrilla and M. S. Devi, "FNU-BiCNN: Fake news and fake URL detection using bi-CNN," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 2, p. 477, 2022.

[54] A. Abbasi, A. R. Javed, C. Chakraborty, J. Nebhen, W. Zehra, and Z. Jalil, "ElStream: An ensemble learning approach for concept drift detection in dynamic social big data stream learning," *IEEE Access*, vol. 9, pp. 66408–66419, 2021.

[55] R. Agrawal, S. Rajagopalan, R. Srikant, and Y. Xu, "Mining newsgroups using networks arising from social behavior," in *Proc. 12th Int. Conf. World Wide Web*, 2003, pp. 529–535.

[56] A. Murakami and R. Raymond, "Support or oppose? Classifying positions in online debates from reply activities and opinion expressions," in *Proc. 23rd Int. Conf. Comput. Linguistics*, 2010, pp. 869–875.

[57] G. Gorrell, E. Kochkina, M. Liakata, A. Aker, A. Zubiaga, K. Bontcheva, and L. Derczynski, "SemEval-2019 task 7: RumourEval, determining rumour veracity and support for rumours," in *Proc. 13th Int. Workshop Semantic Eval.*, 2019, pp. 845–854.

ALAA ALTHENEYAN received the B.Ed. degree in computer and education and the M.S. and Ph.D. degrees in computer science from King Saud University, Riyadh, Saudi Arabia, in 2006, 2012, and 2020, respectively. From 2011 to 2021, she was a Lecturer with King Saud University, where she has been an Assistant Professor with the Computer and Engineering Department, since 2021. Her research interests include natural language processing and machine learning.

ASEEL ALHADLAQ received the B.S. and M.S. degrees in computer science from King Saud University, Riyadh, Saudi Arabia, in 2006 and 2013, respectively, and the Ph.D. degree in computing from Newcastle University, Newcastle upon Tyne, U.K., in 2021.

From 2011 to 2021, she was a Lecturer with King Saud University, where she has been an Assistant Professor with the Computer and Engineering Department, since 2021. Her research interests include human–computer interaction, social media, and designs.

● ● ●