International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Social Media Misinformation

Greeshma R[1], Divyashree Biradar[2], Gargi R Bharadwaj[3], Goutham N[4], Dr. Girijamma HA[5], Mr Sanjay P Kallas[6]

[1,2,3,4,5,6]Department of Computer Science, RNS Institute of Technology, Bangalore,India

Abstract

With over 71 percent of internet users engaged in Online Social Media (OSM), it has emerged as a vital platform for sharing ideas, information, and expressions. However, the credibility of information is not guaranteed due to crowd sourcing and the lack of central moderation. This creates opportunities for malicious users to disseminate rumors and cause panic, especially during real-time incidents or disasters, by generating fake content. Among OSM platforms, Twitter, being a popular micro-blogging website, is particularly vulnerable to the spread of misinformation due to its diverse user base, including the general public, celebrities, politicians, and organizations. This system aims to identify misleading information on Twitter and propose measures that social media companies and users can adopt to prevent the dissemination of misinformation and promote content verification.

Index Terms: Social Media, OSM, vulnerable

I. INTRODUCTION

Digital misinformation, commonly known as online fake news, poses a significant threat to democratic institutions, misguiding the public and potentially inciting radicalization and violence. It typically involves manipulating various media formats such as images, text, audio, and videos to manipulate public opinion, often driven by social, economic, or political motives. Given its pervasive influence on people's beliefs and decisions, there is a pressing need for efforts to detect and combat the spread of fake news, which has become increasingly prevalent in recent years. Detecting fake news on social media presents numerous complex challenges. Unlike other types of content, fake news is intentionally crafted to deceive readers, making it difficult to discern based solely on its substance. It encompasses a wide array of topics, styles, and platforms, aiming to distort facts using various linguistic techniques while mimicking genuine news sources. Existing detection techniques primarily focus on analyzing specific modalities such as text, visual content, or user activities. Although platforms like Politifact, Full Fact, and AltNews endeavor to combat fake news, manual methods are often slow and cannot effectively prevent its initial dissemination. Singlemodality detection approaches are inadequate for identifying falsified content, underscoring the necessity for a multimodal system to effectively detect fake news.

II. UNDERSTANDING MISINFORMATION VS DISINFORMATION

Disinformation is the deliberate spreading of misleading information, while misinformation, as defined by

HimmaKadakas and Ojamets (2022), refers to the dissemination of false or inaccurate information, often unintentionally, without the intent to mislead or deceive the audience. Erku (2021) suggests that misinformation may sometimes be mistaken for disinformation, but disinformation always encompasses misinformation. For example, if a factual error is discovered in an article about a political figure, it qualifies as disinformation (Nikolov, 2020). If the error is found to be intentional, the piece may be labeled as disinformation. While these terms are frequently used interchangeably due to the difficulty in discerning intent, the distinction lies in the purpose of the person or source distributing the information. Misinformation, as per Domenico (2021), does not intend to mislead but rather aims to influence or alter public perception on a specific matter. According to O'Connor and Murphy (2020), "disinformation is false information disseminated with the intent to deceive, whereas misinformation is false information communicated without deliberate malice."

III. RELATED WORKS

Multimodal fake news detection has emerged as a focal point of research interest in recent years. The majority of studies in this field concentrate on analyzing both textual and image characteristics within news articles. Articles displaying deliberately misleading sentiment and containing unverifiable information are typically flagged as potential instances of fake news. Several publicly accessible datasets, including Twitter, Weibo, Gossicop, Politifact, and NewsBag, are commonly employed for research purposes, with Twitter and Weibo being particularly prevalent in this context.

A. EANN: Event Adversarial Neural Network

The Event Adversarial Neural Network (EANN) [2] employs Neural Multimodal techniques to extract text and image features. The FakeNews Detector (Neural Net) determines the authenticity of news articles, while the Event Discriminator filters out specific events from the extracted characteristics. This approach allows for the detection of fake news in new events, enabling EANN to learn event-invariant properties. Achieving an accuracy of 82.7 percent on Twitter and Weibo datasets demonstrates its effectiveness.

B. Introduction of softmax layer

The authors incorporated an attention mechanism to integrate text, image, and social context features. A softmax layer was utilized to distinguish between merged features representing fake and real content. This approach offers the advantage of leveraging social context characteristics like hashtags and emoticons. Furthermore, the attention mechanism facilitates the extraction of relationships between visual elements and the presentation of textual/social data alongside visual features. Achieving a 78.8 percent accuracy rate on Twitter and Weibo datasets, this model demonstrates the effectiveness of its design.

C. Spotfake

Spotfake [4] utilizes Bidirectional Encoder Representations from Transformers (BERT) and the ImageNet Model (VGG) for feature extraction. A Fusion Model (Neural Net) is employed to classify news articles as either real or fake. Notably, it achieves an additional 6 percent accuracy compared to baseline models without the need for an event discriminator or attention mechanism. The model demonstrates an impressive accuracy of 89.2 percent on Twitter and Weibo datasets.

D. LSTM

The authors merge both explicit and latent elements of text and image data into a unified feature space, lev-

eraging these learned features to detect fake news. Compared to a basic Long Short-Term Memory (LSTM)[6] model, the training process is more efficient. The approach integrates latent and explicit image attributes, including resolution, number of faces, text content, number of phrases, and news length. Evaluation on a custom dataset reveals that the model achieves an impressive F1 score of 0.921.

E.Hybrid model

In their work [7], the authors introduce a deep hybrid model designed to learn multimodel correlation embedding. Key components of the model include: (1) Three distinct networks dedicated to processing news images, content, and user profiles; (2) Incorporation of an adversarial mechanism to enforce uniform distribution across various modalities;

(3) Integration of a fully connected neural network-hybrid similarity loss model to capture user sentiment, considering latent feelings. The model aims to create integrated embeddings while considering semantics across multiple modalities. Additionally, it utilizes features such as authors, sources, and keywords. Evaluation on the GossipCop and PolitiFact datasets shows an accuracy of 81.58 percent .

IV. MACHINE LEARNING METHODS TO DETECT FAKE NEWS

To discern fake news, this method predominantly utilized attributes crafted by humans. These attributes included similarities in dissemination structure, geographic location, user influence, and emotional tone, extracted from event-related data through feature engineering. These attributes were then employed to train classifiers such as decision trees (DT), support vector machines (SVM), and others to differentiate between false and genuine news stories (Castillo et al., 2011; Jin et al., 2016; Reis et al., 2019; Wu et al., 2015).Researchers (Castillo et al., 2011) trained a DT algorithm to detect rumors using sentiment scores derived from various criteria, such as the quantity of URLs posted on Weibo and user registration duration. Wu et al. (2015) utilized a support vector machine classifier trained on features such as microblog release location, microblog issuing client, and emotional tone of textual symbols to identify rumors. Reis et al. (2019) proposed a new set of features based on an evaluation of 141 textual features previously suggested for identifying fake news. However, as noted by other authors (Castillo et al., 2011; Mikolov et al., 2013; Popat et al., 2016), successfully crafting hand-crafted features requires expertise in the relevant domain and specific events. Nevertheless, this technique relies on hand-crafted features that lack robustness, and the resulting feature vectors are similarly deficient because the method lacks expertise in fake news detection. Additionally, identifying fake news with custom-built features presents a formidable challenge.

V. DEEP LEARNING BASED MULTIMODALITY MODELS

Numerous researchers have endeavored to automate the generation of deep features through deep learning models to identify fake news. Ma et al. (2016) conducted a study to explore the viability of utilizing deep neural networks for presenting tweets, focusing on temporal and linguistic data collection. Chen et al. (2018) adjusted recurrent neural networks (RNNs) with attentional mechanisms to emphasize various temporal and linguistic aspects. The availability of labeled data is crucial for developing deep learning models, posing a significant challenge historically in the detection of deceptive content. The primary hurdle in rumor identification using deep learning models lies in the complexity of data annotation. To address this challenge, several researchers have explored methods such as unsupervised learning to detect internet

rumors without relying on labeled data. Incorporating a multi-layer recurrent neural network (RNN) into the front end of an autoencoder, as proposed by Chen et al. (2018), notably enhanced the model's performance in rumor detection. A recent study by Raza and Ding (2022) introduced a transformer-based model for detecting fake news, employing an encoder for learning and a decoder for prediction. While the unsupervised learning approach alleviates the need for data labeling, it also introduces inherent instability to the model. Although deep learning's unimodal approach can enhance the accuracy of fake news detection, it overlooks the multidimensional nature of news as a compilation of multimedia data. Fake news holds no value as it disseminates misleading content and images.

VI. METHODOLOGY

A. Data Set

The dataset employed is FakeNewsNet , which stands out for its inclusion of spatiotemporal data, social context, and comprehensive news content compared to other repositories. Each news article within the dataset typically comprises text, images, news titles, and additional metadata. These data originate from two distinct domains: politics and entertainment, sourced from Politifact4 and GossipCop5, respectively.

B. Pre processing and word embeddings

In the pre-processing phase, the data undergoes cleaning to eliminate redundant, unnecessary, or irrelevant information. Text data is prepared for analysis using procedures from the NLTK Python library, including Stop Word Removal, Stemming and Lemmatization, Normalization, and Tokenization. Stop-word removal filters out insignificant words and symbols, while stemming and lemmatization break down sentences into their constituent parts. Normalization ensures that sentences adhere to industry standards, and tokenization divides longer strings into more manageable segments.

Stemming is a heuristic procedure that removes affixes from words to simplify them and facilitate conversion to their base forms. In certain cases, the original word's root form is considered the definitive one. Online content and social media data often contain noise in the form of abbreviations, misspellings, and out-of-vocabulary words, underscoring the importance of text normalization.

To obtain a vector representation of text tokens, a pre-trained GloVe word embedding (Pennington et al., 2014) is utilized. GloVe leverages unsupervised learning techniques to construct word embeddings by summing up the global word-word cooccurrence matrix extracted from a corpus. Compared to other word vector formats, GloVe's use of a co-occurrence matrix to capture global word statistics and meanings offers distinct advantages.

To verify the integrity of multimedia files, the 'Beautiful Soup' Python module (Hajba and Hajba, 2018) is employed to check the links to included images. If any links are broken or images are missing, the associated multimedia files are retrieved from the web. Each instance is then finalized with its three parameters—headline/title, body/text, and image—processed separately for text and multimedia data before being concatenated.

C. Experimental Analysis

The dataset was acquired from Twitter using a web scraping Python script, comprising attributes such as title, text, URL, top image, images, and labels. Preprocessing of the title and text involved removing punctuation, non-ASCII symbols, stop words, alphanumeric words, and links. Image retrieval was

conducted using Python's urllib library. Approximately 8000 and 600 records were gathered from the GossipCop and Politifact datasets, respectively. Two advanced machine learning models, XLNet and VGG19, were employed for experimentation. The title and text data were concatenated and inputted as a single sequence into XLNet, while images were processed using a pre-trained VGG19 model. The output vectors from XLNet and VGG were concatenated with a softmax layer for classification of news items as real or fake. Training was conducted over 100 epochs with a learning rate of 0.0001, utilizing the Adam optimizer and categorical crossentropy as the loss function.

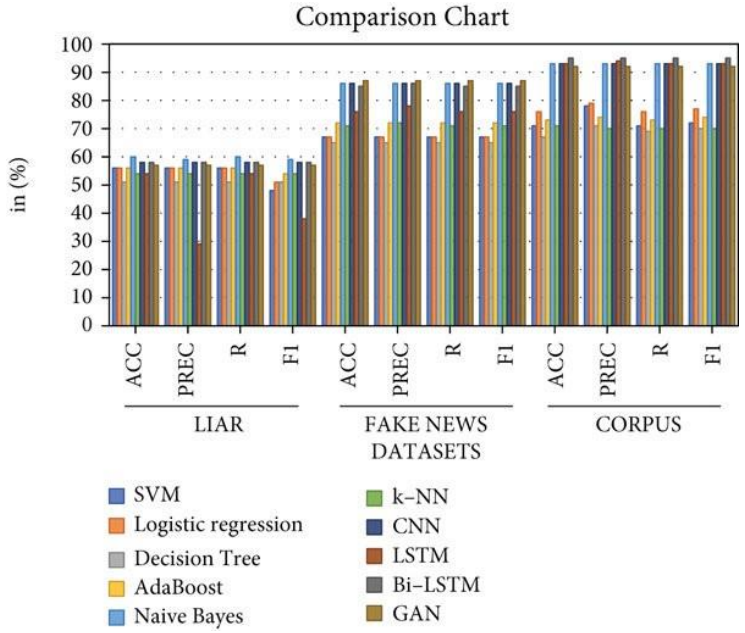
VII. RESULTS

The proposed multi-modal data fusion framework analyzes news titles, textual content, and visual information to identify fake news. Results demonstrate that integrating news titles, textual content, and visual information can enhance the accuracy of fake news classification. The experimental results summary is presented in Table II, which lists the model accuracies of both existing approaches and our proposed method across the datasets. Spotfake+ [12] is a leading model designed for fake news detection, leveraging the FakeNewsNet [10] repository. It boasts remarkable accuracy rates of 0.846 and 0.856 on the Politifact and GossipCop datasets, respectively. Notably, these accuracy levels surpass those achieved by any other multi-modality system utilizing the FakeNewsNet repository. Single-modal fake news detection methods typically achieve accuracies of up to 80 percent. Among the multimodal baseline models, SpotFake+ [12] achieves an accuracy of 85.6 percent on the GossipCop dataset. By incorporating title information into SpotFake+ [12], the model's accuracy improves from 85.6 percent to 87.1 percent on the GossipCop dataset. For training, 4963 records from GossipCop were utilized, with 552 records for validation and 500 records for testing. Figure 2 depicts the plot of training accuracy versus validation accuracy over the number of epochs, illustrating an optimal fit graph. The model achieves an accuracy of 87 percent over 100 epochs. Figure 3 displays the graph of training loss versus validation loss over the number of epochs. Both training and validation losses decrease until reaching a point of stability, with a minimal gap between the curves indicating a well-fitted model.

VIII. CONCLUSION

A sophisticated multimodal fake news detection model has been developed. The approach combines three distinct features of a news item: title, textual content, and visual information. Specifically, the news title and textual data are merged and pre-trained using XLNet, while visual features are extracted through the VGG-19 model. The resulting feature vectors

Fig. 1. Bar Graph



"Spotfake: A multi-modal framework for fake news detection," in 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM), 2019, pp. 39–47.. K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media," Big Data, vol. 8, pp. 171–188, 06 2020 S. Jindal, M. Vatsa, and R. Singh, "Newsbag: a benchmark dataset for from XLNet and VGG19 are concatenated to make the final prediction. This model has demonstrated an accuracy of 87 percent on the GossipCop dataset. Moving forward, the plan is to further enhance the approach by incorporating features from multiple images.

TABLE II
COMPARISON OF ACCURACY BETWEEN BASELINE AND PROPOSED MODELS

Modality	Models	Politifact	Gossicop
Text	SVM	0.58	0.497
	Logistic Regression	0.642	0.648
	Naive Bayes	0.617	0.624
	CNN	0.629	0.723
	SAF	0.691	0.689
	XLNET + Dense Layer	0.74	0.836
	XLNET + CNN	0.721	0.84
	XLNET + LSTM	0.721	0.807
Image	VGG19	0.654	0.80
Multimodal (Text+Image)	EANN [2]	0.74	0.86
	MVAE [1]	0.673	0.775
	SpotFake [4]	0.721	0.807
	SpotFake+ [12]	0.0.846	0.856
	SpotFake+ updated	**0.818**	**0.871**

REFERENCES

1. D. Khattar, J. S. Goud, M. Gupta, and V. Varma, "Mvae: Multimodal variational autoencoder for fake news detection," in The World Wide Web Conference, ser. WWW '19. New York, NY, USA: ACM, 2019, p. 2915–2921. [Online]. Available: https://doi.org/10.1145/3308558.3313552
2. S. Singhal, R. R. Shah, T. Chakraborty, P. Kumaraguru, and S. Satoh, fake news detection," 2019
3. S. Singhal, A. Kabra, M. Sharma, R. R. Shah, T. Chakraborty, and P. Kumaraguru, "Spotfake+: A multimodal framework for fake news detection via transfer learning," 05 2020.
4. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
5. G. Krishnamurthy, N. Majumder, S. Poria, and E. Cambria, "A deep learning approach for multimodal deception detection," 03 2018 Fig. 2. Comparison