IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 2, February 2023

DOI: 10.17148/IJARCCE.2023.12242

"Fake News Detection Using Machine learning"

Mrs. Usha. M[1],

Lakshmi N S[2], Lakshmi N[3], Divyashree P[4], Deenakumari K[5]

Department of Computer Science and Engineering, East West Institute of Technology, Bangalore, India[1]

Department of Computer Science and Engineering, East West Institute of Technology, Bangalore, India[2-5]

Abstract: False information concerning topics or events, such as COVID-19, is referred to as fake news. Social media juggernauts claimed to take COVID-19-related falsehoods seriously at the same time, yet they were ineffective. Real news data are gathered for this study via information fusion from websites related to news broadcasting, health, and the government, whereas false news data are gathered from social media platforms. Using cutting-edge deep learning models, 39 features were extracted from multimedia texts and utilised to identify bogus news about COVID-19. The accuracy of our model's false news feature extraction increased from 59.20 to 86.12 percent. Our best recall and F1-Measure for fake news were 83% utilising the Gated Recurrent Units (GRU) model, which has an overall high precision of 85%. Similarly, F1-Measure for real, recall, and precision

Keywords: Fake news, Social media, Deep learning, NLP, Mining Emotions

I. INTRODUCTION

The existence of "Fake News" material is nothing new. With Despite this, they have quickly spread thanks to the internet. The COVID-19 epidemic has highlighted the limitations of modern technologies and the Internet of Things (IoT), which foster a number of conspiracies and lead to people utilizing fake corona virus treatments or making misleading declarations. These dangerous health suggestions caused a number of people to become exposed to much worse diseases than the virus itself. Lawmakers, industry participants, and the general public have displayed confusion, apprehension, and suspicion.

Enemies have been formed, and some citizens have disobeyed orders to stay at home from the administration. Evidence gathered from multiple multimedia platforms, including Facebook, shows that in April 2020, about 50 million pieces of misinformation and deception relating to COVID-19 were removed.. the reverse side, Twitter questioned over 1,500,000 of its people who disseminating misleading information and engaging in what they referred to as false "manipulative behaviour" within the same time frame. While many published movies on YouTube with false material on the corona virus were taken down, around 18 million scam emails were recognized and stopped in Gail and YouTube.

The writers of agree the fact that multimedia services developed into a bastion of slanderers propagating false information, uncensored news source using the intent to cheat, mislead, and confuse other users. much information is currently available about bogus news on the internet and the bizarre situations produced for internet users. Nobody will contest the existence of social media usage rose as a result of inactivity and the policy of staying at home, supporting the author's argument that fake news exploded and offered a framework to address it in.This method integrates the pixel and frequency branch and uses visual data from them to identify bogus news in the network Convolution neural network-recurrent neural network [CNN-RN] model with several branches is utilized to extract the collected picture features by the publishers, who designed using a CNN-based network Further research on a set of real-world data showed that they use outperforms other current state-of-the-art algorithms with precision improvement of 9.2% and an improvement in performance detection of above 5.2%.

The contributions of this study are as follows:
• The study collects actual news data from numerous multimedia platforms using the Information Fusion process: a New York

Global Health Today, Times, Harvard Health, the CDC, and the WHO are all organizations that focus on preventing and controlling disease (WHO) (CDC). notwithstanding the fact that for gathering of bogus news from YouTube, FB, and other media platforms.

• Sets of data utilized solely includes class labels for text and,it performs poorly recurrent classifiers for neural networks, LSTMs (long short-term memories), and neural networks [RNN] [GRU]. We have suggested 39 novel text features to

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 2, February 2023

DOI: 10.17148/IJARCCE.2023.12242

enhance Recall, accuracy, precision, and F1-Measure. To our knowledge, no such features have ever been employed for fake news.

• We suggested named entity-based features, sentiment features, and linguistic features. Our updated features have an accuracy of 86.12 in identifying COVID-19 bogus news.once the text's features have been extracted. Hence, new capabilities result in a 20% increase in accuracy.

the remaining this materials set up in an ordered as in the following : Part 2 provides a thorough explanation of the relevant studies and a broad perspective on fake currency identification news with their prevention techniques. The research methodology is explained in the experimental setup is provided in the evaluation work, results, and performance analysis are described in and contains a conclusion from the work.

II. CONNECTED WORKS

The differences between President Trump's and Prime Minister Trudeau's attitudes towards the COVID-19 epidemic are evident through a quantitative analysis of the themes that appeared on each person's Twitter feeds during the height of the pandemic. One's emerging theme is politics, while the other has been public health and policy . In order In order Bonito and Nazareth use Network Science, which considers the relationships between systems in the formation of "co-occurrence networks," to connect the terms that are present in the two tweets. 'COVID-19' and 'pandemic,' for example, were both mentioned in the same tweet.. Based on how frequently they appeared, @realDonaldTrump and @Justin Trudeau listed their top 100 phrases. appeared retrieved. According to the network analysis, politicians' social media posts influence how the public perceives the pandemic and how efforts to control it should be made in light of those perceptions.Once again, significant improvements in sentiment analysis jobs have been made as a result of the development of numerous deep neural network models. For sentiment analysis tasks, these neural network models, however, were unable to precisely capture sentiment information, resulting in their instability. To create the final sentiment, aspect-category or aspect-terms have been merged. The feelings and attitudes of software developers have a big impact on the output and calibre of their work. Current efforts on sentiment analysis and emotion mining are still when it comes to accuracy, datasets size, and analytical capabilities, software engineering is in its infancy specialization. Incorrect information about the COVID-19 pandemic is a critical issue that may have a significant impact on how people react to the virus. Individuals are urged to double-check information regarding the epidemic from a variety of multimedia sources, including the WHO's official website, by reading the entire piece rather than just the headline, and by looking up the author's credentials. Recent research has revealed that visual material, such as pictures and videos, nearly always goes along with bogus news and can be very helpful in determining which is phoney.

III. PROPOSED ARCHITECTURE

Using cutting-edge deep learning models, 39 characteristics were taken from texts and used in this study to spot fake news about Covid-19 on social media.

3.1. Datasets summary

The COVID-19 datasets that was used in this study was put together from [25]. Almost 1100 news items are included in the collection COVID-19 is the subject of articles and media pieces, 586 genuine news stories and 578 fraudulent news stories. The CDC, WHO, The New York Times, and other reputable sources were used to compile the actual story. False news was obtained from medical websites and social media posts on Facebook.

3.2. Text feature extraction

The process of feature extraction involves a dimensionality reduction in which an Condensed versions of the original raw data collection are divided into more manageable processing classes. By doing this, we decrease the processing time of the compiler and enhance the word value detection efficiency. Such large data sets have the drawback that some variables require a lot of computational power to process. We have therefore developed our own text characteristics to identify bogus news for this reason. These characteristics include the number of stop words, the percentage of capital letters, small letters, and capital letters, the amount of numerical values, the word and character counts, the number of sentences and the texts' average sentence length. Also, it contains the named entity recognition (NER) characteristics that were extracted from text, together with the emotion scores, Average word count, sentence polarity, and positive, negative, and neutral sentiment negative, neutral, and complicated. From the sentences, we retrieved features, and these features served as the basis for classification.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 2, February 2023

DOI: 10.17148/IJARCCE.2023.12242

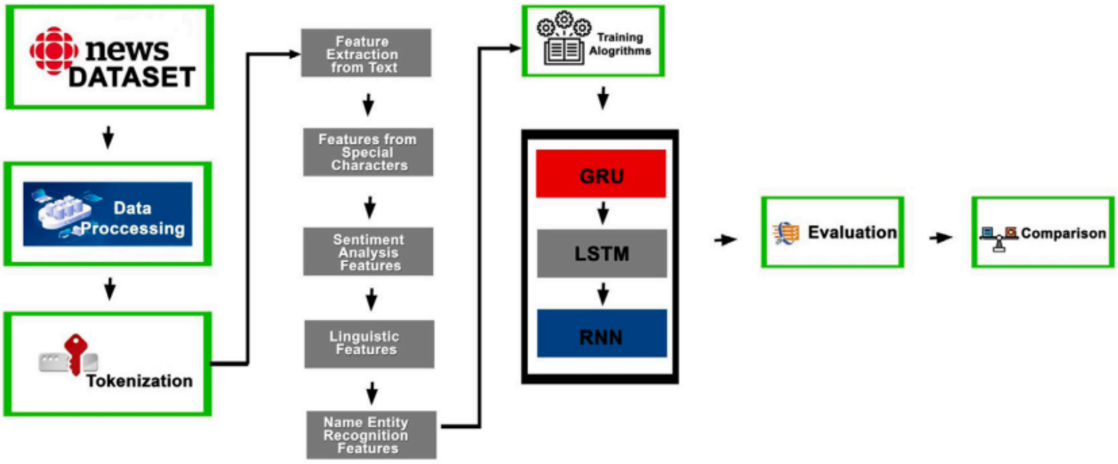

Fig. 1. Proposed model.

3.3. Special character characteristics

Special character traits depict the scenario that occurs after data gathering. According to our approach, each input vector has a distinctive feature in it. For instance, To convey a variety of characters, min-char-inn uses one-hot vectors. Using the previous character to predict the subsequent one one, it enables our algorithm to give real training phases the preceding character. Shorter text classification is made possible by using a particular character that combines fully connected layers of RNN, LSTM, and GRU.

3.4 . Sentiment analysis capabilities

It is essential to categorise positive, negative, neutral, or complex sentiment in our model texts or sentences before applying the sentiment analysis features. Our generated text in this case has a numeric data type. It displays the characteristics we produced from the texts and applied to categorise.

3.5. Language characteristics

The classification process used in this study involves assigning a category to a piece of fake news material. This is accomplished by translating that any other language into text. The attribute names are organised according to their data kinds, showing both numerical and non-numerical traits.

3.6. Features for named entity recognition, section

The terms These terms also include entity extraction, entity identification and entity chunking. Text tokens are intended to be assigned and classified using Named Entity Recognition (NER) into specified groups.

IV. RESULTS AND ANALYSIS

These tests are all carried out using Google Colab Core I3 processor and 8 GB of Memory and a 2.7 GHz processor make up the system specs.

4.1 Findings that do not extract any features, section

We can demonstrate that the Ada-boost classifier performed better than other machine learning techniques like Decision Tree (DT) and K Nearest Neighbor in terms of the F1-Measure score, recall, accuracy, and precision . The highest prediction precision of any machine learning classifier, 79.88%, is attained using Ada-boost prior to feature extraction. Comparable results were obtained with Ada-boost, including , 81.82% F1-Measure score, 86.36% recall, and 76.76% precision. For DT and KNN, prediction and accuracy are 67.81% and 62.06%, respectively. For DT, the corresponding values are 70.51%, 62.50%, and 66.26% for F1-Measure, recall, and precision, respectively. KNN obtained a precision of 72.91%, a recall of 39.77%, and51.47% F1-Measure in each case. We can see that, prior to developing our own features as described in Section 3, the conventional machine learning algorithms outperformed our suggested deep learning models quite well. Consequently, it is anticipated that our suggested model will perform better than machine learning models once we have created our innovative features from text.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 2, February 2023

DOI: 10.17148/IJARCCE.2023.12242

The F1-Score, Precision, and Recall for without feature extraction, the GRU model seen to be 55%, 58%, and 56% for false news and 63%, 60%, and 62% for actual news, respectively. when employing the Model GRU. Similar to this, the GRU model's training loss is 0.05 and its training accuracy is 98.29%. Due to the data's refinement, it is common to anticipate that an accuracy test will to be less than a training precision. Furthermore, forecast Scores for a loss of forecast are 4.11 and 59.20% accurate, respectively We managed to attain 6.90 testing loss and 0.31 training loss. using the LSTM model. For the LSTM model, the accuracy during training and testing is 95.07% and 55.72%, respectively. Using an LSTM model, the accuracy, and F1-Measure for recall bogus 51%, 60%, and 55%, respectively, of news. The F1-Measure, recall, and precision scores actual news are also 56, 52, and 61 percent, respectively. For the training loss for RNN models is represented as 0.00 and testing loss as 3.30. The RNN model's Accuracy during training and prediction is 100% and 57.71%, respectively. loss and accuracy when Comparing training on GRU and LSTM, the RNN general model obtained zero training loss 100% training precision. Similar to this, RNN accomplished minimal loss for prediction, 3.30 and GRU attained high forecast perfection.

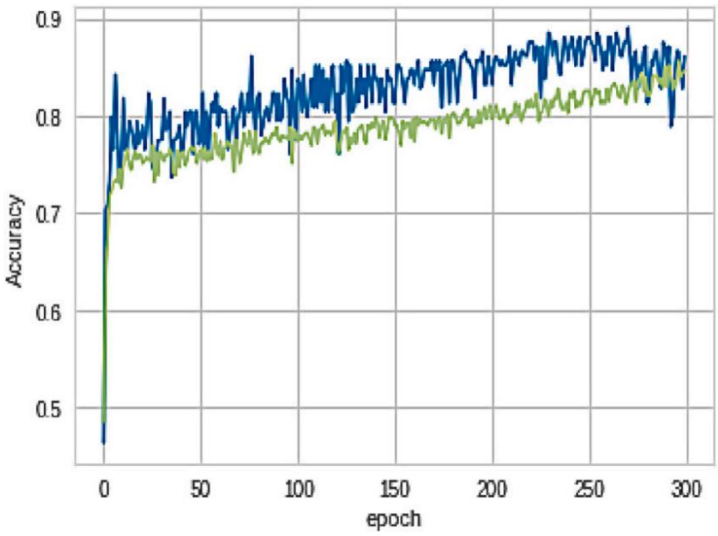

Fig. 2. GRU model training and testing accuracy.

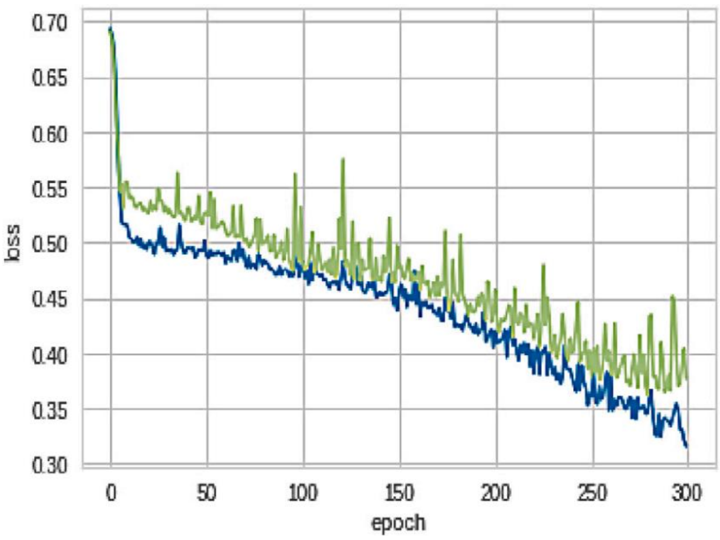

Fig. 3. GRU model training and testing loss

compared to LSTM and RNN, 59.20% High precision for bogus news using the GRU model is 55%. With an RNN model, the optimum memory and F1-Measure for bogus news were 62% and 57%, respectively. For genuine Currently, 63%, 60%, and 62% are the optimal precision, recall, and F1-Measure values in GRU mode, respectively. Because just 30% of the datasets is used for testing and 70% is used for training, training accuracy is excellent. Additionally, it demonstrates how well deep learning algorithms work with enormous datasets and characteristics. illustrates the GRU model's training

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁘ Impact Factor 7.918 ⁘ Vol. 12, Issue 2, February 2023
DOI: 10.17148/IJARCCE.2023.12242

and prediction accuracy. The training accuracy is shown by that curve in green. The highest training accuracy is 98.29% at the 100th epoch, and it ranges from 45% at the first epoch. Likewise, the LSTM forecast for the first epoch The highest reported prediction accuracy after the 100th epoch is 55.72% for the LSTM model, which has a prediction accuracy of 56%. Displays loss of the LSTM model in training and testing.

This time, the training loss is represented by the blue curve, and the testing loss is represented by the green curve. The training loss is roughly 0.68 at the first epoch and is 0.00 by the 100th iteration. demonstrates the Model RNN's prediction and training accuracy. Similar to this, prediction loss at the first epoch is 0.67 and prediction loss after the 100th epoch is 0; in the first epoch, it is 53.32%, and we attained 100% at the 100th epoch, making this the best training accuracy. The same was true for the first epoch's forecast accuracy utilizing After the 10th epoch, the greatest recorded prediction accuracy for the model is 57.71%. RNN model loss is shown both for testing and instruction. Loss of training is approximately 0.70 at the first epoch and 0.58 after 10 iterations. With the GRU model, the prediction loss at the first epoch is 0.72 and at the tenth epoch is 0.68, respectively. presents the AUC splines for the classifier GRU, with an AUC of about 59.05%. similarly displays the LSTM AUC curve and AUC of the RNN model. The RNN and RNN AUCs are LSTM are Both 56.11% and 57.99% are acceptable. When compared to other classifiers, GRU performed better in terms of AUC.

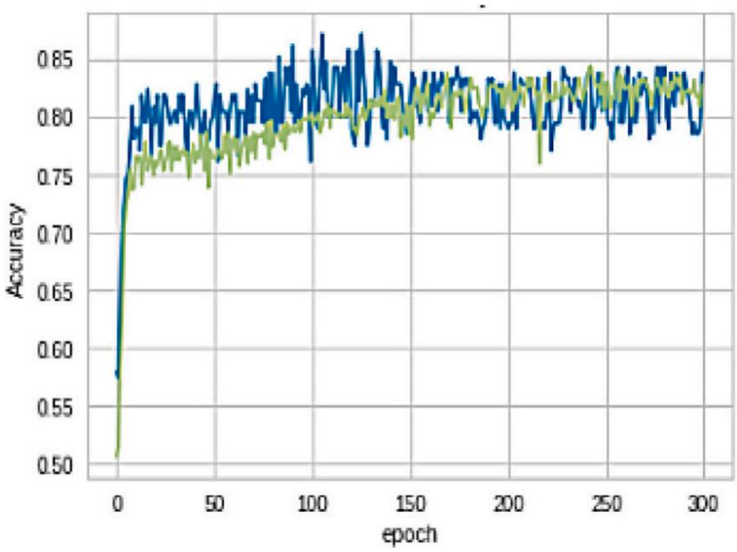

Fig 4. LSTM model training and testing accuracy.

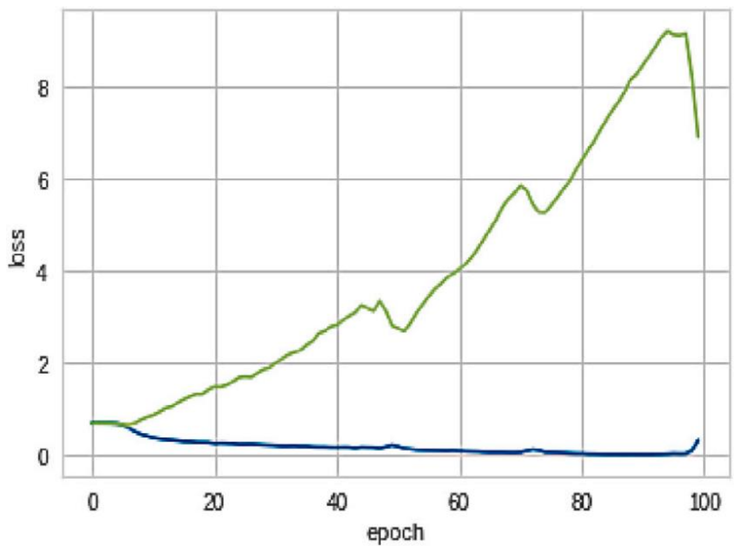

Fig 5. LSTM model training and testing loss.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 2, February 2023

DOI: 10.17148/IJARCCE.2023.12242

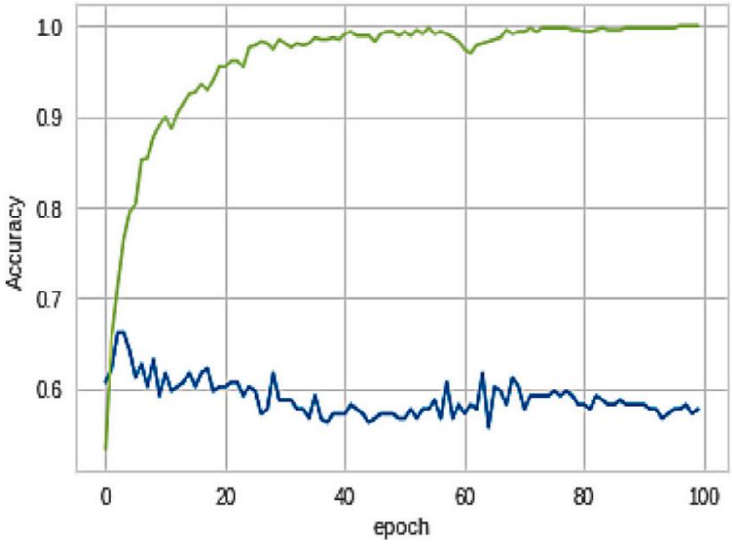

Fig 6. RNN model training and testing accuracy.

4.2. Output from the features extraction

According to this, the Ada-boost classifier had a prediction accuracy score of 82.75%, while the The results for the DT and KNN had respective values of 77.58% and 69.54%. Likewise, Precision, Recall, and F1-Measure Values are 79% and 89.77%, respectively all scored at 84.04%.With DT, one may attain scores for F1-Measure, recall, and precision were 72.47%, 89.77%, and 80.20%, respectively. The KNN classifier's recall, accuracy, and F1-Measure score were each 62.41%, 100%, and 76.85%, respectively. When both DL and ML features had been retrieved, it was discovered that deep learning approaches performed better than machine learning ones in terms of precision, recall, F1-Measure, and prediction accuracy. Accuracy and Loss for each classifier Before feature extraction, this displays the Classification Report.

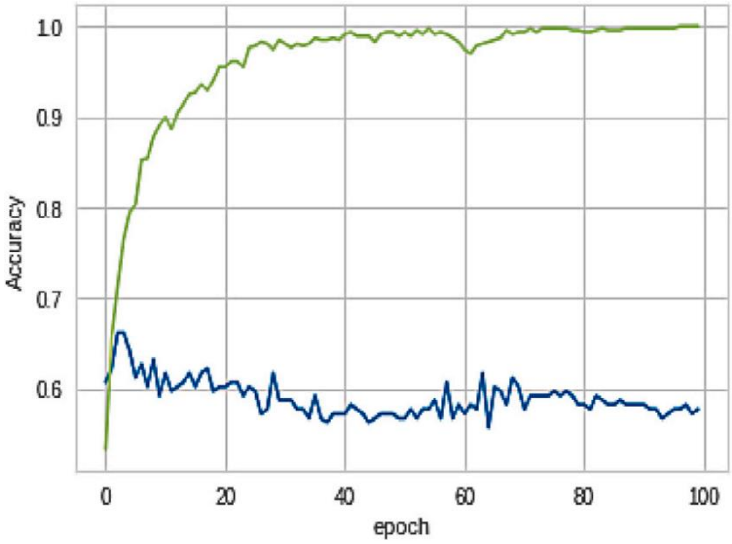

Fig. 7. RNN model testing and training failure.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 2, February 2023
DOI: 10.17148/IJARCCE.2023.12242

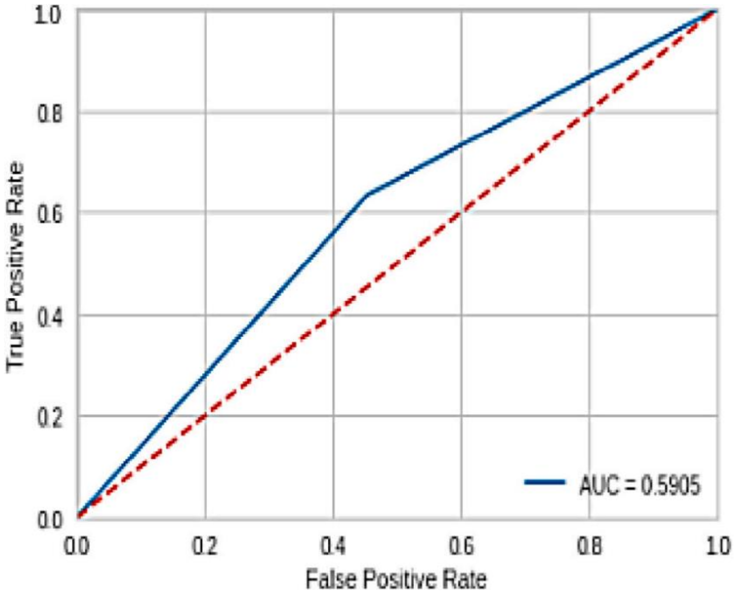

Fig. 8. GRU ROC curve.

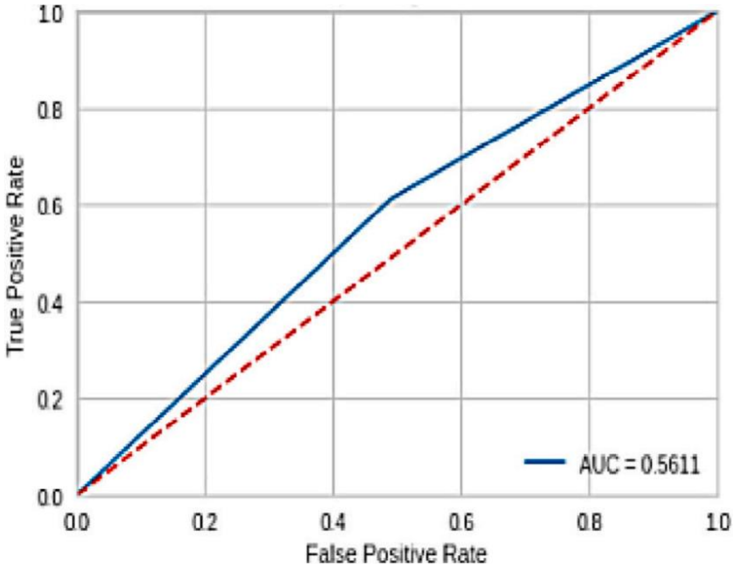

Fig. 9. LSTM ROC curve

V. CONCLUSION

As a result of the COVID-19 pandemic, social isolation rules and mask use have the potential to significantly reduce or even end the coronavirus spread. Why is it so difficult to convince individuals to use these straightforward strategies? There are many causes, but this study has highlighted internet misinformation as a key one. As a result, people are quite unsure and find it difficult to comprehend the virus's exponential spread. An spread of false information regarding the virus coincided with the infection's outbreak. This incorrect information includes false claims about the origin of the virus, conspiracies, phoney treatments, and hazardous medical advice. Misinformation puts public health interventions in danger.

This study introduced a revolutionary system with 39 properties in spotting false information about COVID-19. The model applies cutting-edge deep learning models includingGRU, LSTM, and RNN and uses an information fusion approach to acquire social media data. We have suggested named entity-based features, sentiment features, and linguistic features. Our new features, which extract features from the text, have an accuracy of 86.12% in identifying false information on COVID-19. As a result, accuracy is up 20%. With an RNN model, the overall high precision is 85%. The

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⸎ Impact Factor 7.918 ⸎ Vol. 12, Issue 2, February 2023

DOI: 10.17148/IJARCCE.2023.12242

outcomes also demonstrate that 83% using GRU is the best recall and F1-Measure for bogus news.model. Similarly, the accuracy, For true news,Recall and F1-Measure had respective values of 88%, 90%, and 88%. For same datasets, our solutions performed better than the conventional machine learning algorithms.

REFERENCES

[1]Petchot Kamonwan. ‘Fake news’ in the time of COVID-19. 2020, UNESCO Bangkok. URL https://bangkok.unesco.org/index.php/content/press-provides antidote -fake-news-time-covid-19. [Accessed 10 July 2020].

[2] Bedi A, Pandey N, Khatri SK. A framework to identify and secure the issues of fake news and rumours in social networking. In: 2019 2nd international conference on power energy, environment and intelligent control. 2019. p. 70–3.

[3] Qi P, Cao J, Yang T, Guo J, Li J. Exploiting multi-domain visual information for fake news detection. In: 2019 IEEE international conference on data mining. 2019. p. 518–27.

[4] Bonato A, Nazareth A. Facts or fake news: Revealing patterns in the COVID-19 tweets of Trudeau and Trump. 2020, URL https://theconversation.com/factsor-fake-news-revealing-patterns-in-the-covid-19-tweets-of-trudeau-and-trump-142139. [Accessed 10 July 2020].

[5] Ding Jin, Sun Hailong, Wang Xu, Liu Xudong. Entity-level sentiment analysis of issue comments. In: 2018 IEEE/ACM 3rd international workshop on emotion awareness in software engineering. 2018. p. 7–13.

[6] Snell N, Fleck W, Traylor T, Straub J. Manually classified real and fake news articles. In: 2019 international conference on computational science and computational intelligence. 2019. p. 1405–7.

[7] Uppal A, Sachdeva V, Sharma S. Fake news detection using discourse segment structure analysis. In: 2020 10th international conference on cloud computing, data science engineering. 2020. p. 751–6.

[8] Manzoor SI, Singla J, Nikita. Fake news detection using machine learning approaches: A systematic review. In: 2019 3rd international conference on trends in electronics and informatics. 2019. p. 230–4.

[9] Ramezani M, Rafiei M, Omranpour S, Rabiee HR. News labeling as early as possible: Real or fake? In: 2019 IEEE/ACM international conference on advances in social networks analysis and mining. 2019. p. 536–7.

[10] Vinit Bhoir S. An efficient fake news detector. In: 2020 international conference on computer communication and informatics. 2020. p. 1–9.

[11] Sa-nga-ngam P, Mayakul T, Srisawat W, Kiattisin S. Fake news and online disinformation. A perspectives of Thai government officials. In: 2019 4th technology innovation management and engineering science international conference, TIMES-ICON. 2019. p. 1–4.

[12] Barua R, Maity R, Minj D, Barua T, Layek AK. F-NAD: An application for fake news article detection using machine learning techniques. In: 2019 IEEE Bombay section signature conference. 2019. p. 1–6.

[13] Das D, Clark AJ. Satire vs fake news: You can tell by the way they say it. In: 2019 first international conference on ?transdisciplinary AI. 2019. p. 22–6.

[14] Kang S, Hwang J, Yu H. Multi-modal component embedding for fake news detection. In: 2020 14th international conference on ubiquitous information management and communication. 2020. p. 1–6