



An Advanced Hybrid Deep Learning-Based Early Rumor Detection Model for Social Media Users

Reeta Bourasi¹, Dr. Sakshi Rai²

^{1,2} Department of Computer Science and Technology, LNCT University, Bhopal, Madhya Pradesh India

Abstract. The importance of social media posts has been increased in our daily life activities. The wide distribution of different rumors and misinformation has affected many business organizations. The wide spread of rumors on social networking sites has brought many social, economic, personal, and political problems throughout the world. False information distorts public perceptions, undermines corporate reputations, disrupts communities, diminishes governmental and business integrity, and poses a significant threat to societal stability. False information affects public perception, harms corporate reputations, and threatens societal stability. Previous machine learning methods like Random-Forest and Gradient-Boosting have been used to detect rumors. However, the proposed hybrid model employs a Convolutional Neural Network combined with Bidirectional Long Short-Term Memory and Gated Recurrent Unit (CNN-BiLSTM and CNN-BiGRU) for classifying social media posts. This paper proposes an advanced hybrid deep learning-based approach that targets early rumor detection on Twitter and Facebook, achieving an accuracy of 98% - 99%, which surpasses baseline machine learning models.

Keywords: Mental-health, Rumor Detection; Deep Learning; Machine Learning; Social Network and Social-media-post-analysis.

I. Introduction

A considerable portion of persons in modern society relies on social media websites like Facebook, Twitter, Snapchat, YouTube, and Instagram to create and nurture interpersonal relationships. Suicide, the most severe acute outcome in psychiatry, results in approximately one million fatalities globally each year. This survey paper reviews current deep learning research focused on rumor detection and prevention, aiming to identify distinctions among these studies, research gaps, and challenges. It discusses advanced methods of rumor detection, their performance on benchmark datasets and approaches such as rumor verification and stance classification. In addition, a brief description of recently used datasets and their research gap is discussed for advancing rumor control techniques [1].

The availability of Internet has drastically increased and so the use of social network has increased which has millions of users in it. But, there has been pollution of information due to which rumors are widely spread. This paper focuses on the

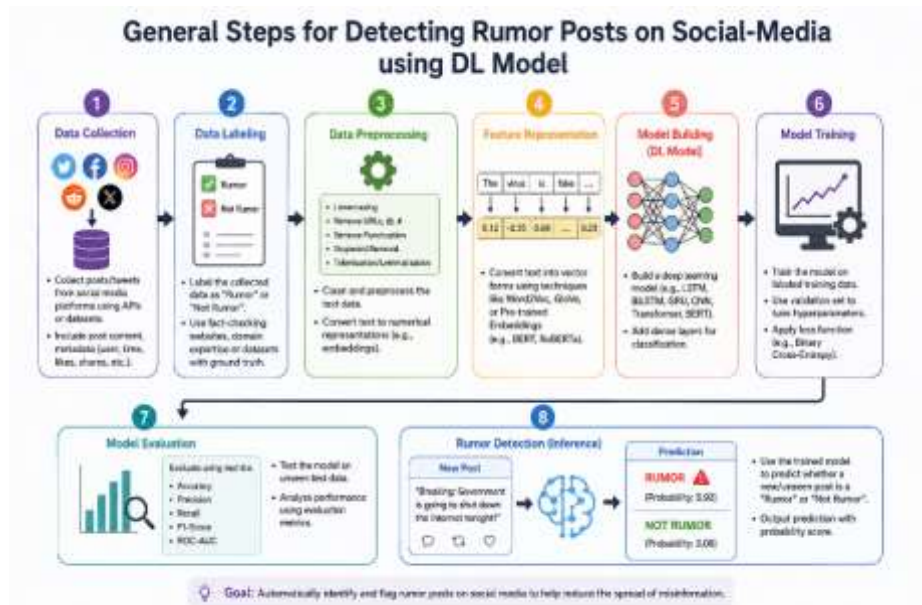


Fig. 1 : General Steps for Detecting Rumor Posts on Social-Media using DL Model

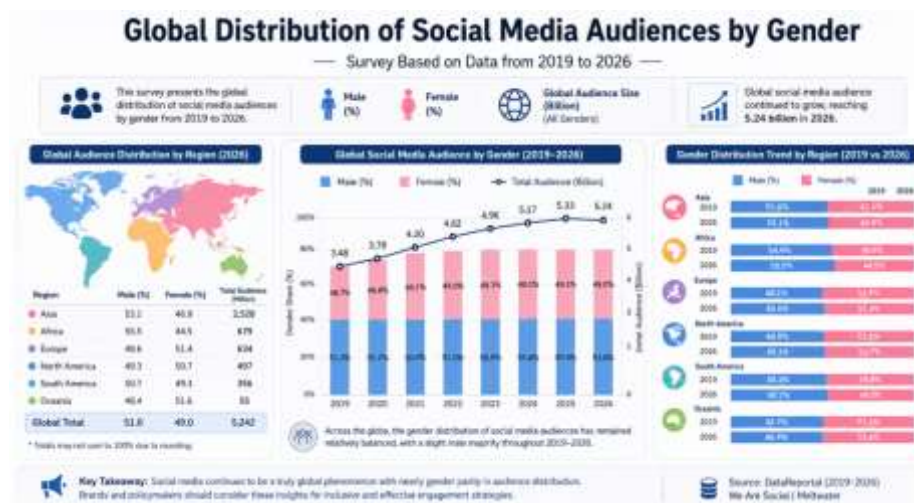


Fig. 2 : The Globally Distributed Social Media Users Datasets by their Gender based as of 2019 to 2026 latest detection methods, categorization of rumor propagation models and challenges and future scope in this field [2].

Social media and micro-blogging platforms help in rapid sharing of information but also help in the fast spreading of rumors which may cause chaos. The existing techniques of early rumor detection depend upon the content and user profiles but do not take into account the importance of propagation patterns of rumors [3]. Figure 1 illustrates the ongoing procedure for detecting rumour posts behaviours in social media comments utilizing a deep learning model, specifically involving dataset preparation, and processing initiation of the processed dataset to the training model [4].



II. Review Literature

Several researches are being conducted on how to tackle the problem of rumour detection. Figure 2 illustrates the study of social media audience based on gender through survey conducted from 2019 to 2026. One of the current issues is the problem faced by organizations in detecting the fake news online. It is difficult to differentiate the deceiving information; deep learning has proven to be more accurate than conventional machine learning. In contrast to the prior review of the literature that mostly concentrated on data mining, this research focuses on advanced deep learning which has been ignored in the previous surveys. It describes the implications of the fake news, reviews datasets and NLP from prior researches, classifies deep learning approaches and evaluation metric. Moreover, it makes recommendations on improvement of fake news detection in future researches [5]. Mental health can be defined as the overall well-being of a person emotionally, socially and psychologically. It affects one's capacity to think, behave and feel. The mental health of a person involves his/her ability to cope up with stress and make sound decisions [6-7].

Nonetheless, preserving the-integrity-of-such information remains a persistent difficulty [7]. A-recent-study conducted by MIT-researchers indicates that misinformation disseminates significantly more rapidly than factual news. Historically, rumours concerning stock markets, including corporate announcements, earnings forecasts, and assessments of inexpensive stocks, have influenced share prices [8-9]. Facebook and Twitter were heavily used by candidates and supporters to advertise and communicate their feelings during the 2017 US Constitutional Election. 529 different stories about Hillary Clinton and Donald Trump surfaced in the media during the event [10-12]. Through social media, the stories reached thousands of followers and had an impact on the election [13-17]. The press's Twitter account was suspended in 2013 for spreading misleading information about two White House incidents surrounding the president's ailment [18-21].

This review also looks at the common concerns found in IoT-enabled wearable sensor networks, the specific issues that need to be addressed in order to maximise the profit potential of these systems in healthcare, and future architectural and technological changes that can be implemented to improve the entire industry [22-23].

The research work provides a novel approach for the effective detection of rumours. The suggested model utilizes the CNN-BiLSTM-BiGRU methodology to effectively classify posts into rumour and non-rumour categories. The primary contribution of this research study can be summarized as follows:

- i. A thorough literature review examining diverse methods, approaches, and various machine and deep learning algorithms devised by researchers to address the issue of



rumour detection. and conduct a gap analysis based on a literature review to provide an improved solution.

ii. Design and Analysis of results from several advanced machine learning methods, including an advanced hybrid deep learning-based termed CNN+BiLSTM+BiGRU, utilized for early rumour identification across diverse datasets.

Table I. Comparisons Of ML And DL Algorithms Utilized By Researchers For Rumour-Detection System

Ref.	Platform	Algorithm/Technique	Features	Results	Drawbacks
[9]	Twitter (X)	SVM, J48, Naive Bayes	Content, user and social features	89% accuracy	Labelling needs to be done manually
[10]	Twitter (X)	CNN-BiLSTM	Content, network, Twitter-specific features	92% accuracy	Only-single-feature TF-IDF was used
[11]	Twitter (X)	J48, DT	Content, network, Twitter-specific features	82 % F-Measure	Work was restricted for implementation only by WEKA
[12]	Twitter (X)	Clustering-method	Content features	90% Precision	Response time was slow
[13]	PHEME	RF, SVM, NB	Content, user features	74.6% accuracy	Few features were used
[14]	Twitter (X)	RNN	Word2Vec	73% accuracy	Accuracy needs to improved
[15]	PHEME	BiLSTM-CNN	Word2Vector and FastText	86.12% accuracy	Accuracy needs improvement
[16]	PHEME and Twitter PHEME	LSTM, LSTM- CNN	Twitter based features as well as metadata	68% accuracy	High Cost of computation
[17]	PHEME and Twitter PHEME	LSTM, LSTM- CNN	LSTM, LSTM with dropout, LSTM-CNN	82% accuracy	Performance needs improvement

III. Proposed Model

This work presents a novel methodology that accurately detects rumors, accommodates numerous datasets from various platforms, and reduces the time required for rumor identification. The proposed hybrid model (also shown in Figure 3 as proposed framework) possesses the following key features:

- 1. Enhanced accuracy :** The suggested model demonstrates superior rumor detection accuracy relative to other baseline models.
- 2. Automatic rumor detection :** Our proposed model autonomously identifies rumors at preliminary stages without human intervention.
- 3. Automated feature detection :** Feature detection is a critical criterion for the development of any model. The suggested hybrid model autonomously and effectively identifies characteristics from the datasets.



Feature-Extraction

Feature extraction is conducted subsequent to data preprocessing. In the feature extraction procedure, the integer vector of the sentence is converted into a dense vector. Machine and deep learning models cannot directly process text; therefore, it is necessary to translate the text into numerical form. Embedding is utilized to transform text into numerical representations.

Algorithm 1 Data Preparation_Algorithm

Input: $T_w \leftarrow$ Tweets set , $L_a \leftarrow$ labels on dataset ,

$L_m \leftarrow$ maximum words used in a tweet

Output: $E \leftarrow$ Embedding matrix

```
1 for t in  $T_w$  do
2    $T_w \leftarrow \text{filterTwitterMarkers}(tw)$ 
    $T_i \leftarrow \text{filterIrrelevant}(T_w)$ 
    $T_d \leftarrow \text{filterDuplicate}(T_i)$ 
    $T_c \leftarrow \text{convertCase}(T_d)$ 
3 End
Encoding:
    $\text{toke} \leftarrow \text{Tokenizer}()$ 
    $\text{toke.fit\_on\_texts}(T_x)$ 
    $T_{xs} \leftarrow \text{toke.texts\_to\_sequence}(T_x)$ 
    $w_i \leftarrow \text{toke.word\_index}$ 
    $\text{wordcount} \leftarrow \text{len}(\text{word\_index})$ 
    $I \leftarrow \text{labelEncoder}()$ 
    $L_e \leftarrow I.\text{fit\_transform}(L)$ 
Padding:
    $T_p \leftarrow \text{pad\_Sequences}(T_{xs}, I)$ 
Custom Embedding Matrix:
    $C \leftarrow \text{Custom}()$ 
4 for word, index in  $\text{word\_index.items}()$  do
5    $W \leftarrow C.\text{get}(\text{word})$ 
    $E[\text{index}] \leftarrow W$ 
6 End
```

Various strategies exist for generating word embeddings for deep learning models, including one-hot encoding, TF-IDF, Word2Vec, custom embeddings, and GloVe embeddings. In the proposed research, we employ the TF-IDF and CountVectorizer feature extraction methods for machine learning, together with unique embeddings to capture the semantic relationships among the text while applying deep learning algorithms. Term frequency is a weighting matrix utilized to assess the importance of a word (count



+ weight) within a document relative to a dataset. TF-IDF is a composite of two concepts: term frequency (tf) and inverse document frequency (idf). Technical steps and process of proposed algorithm has been described in algorithm 1 (data preprocessing and algorithm 2 (proposed model for dataset).

After the datasets are cleaned and feature-extraction is accomplished, we deployed multiple advanced ML and DL algorithms, and their performance was evaluated on the datasets [17-18]. Subsequent to machine learning, multiple deep learning techniques were applied to the dataset [19].

Algorithm 2 Proposed Model

Input: E, W

Output: classification of tweets

Creation of Model:

```
Model ← Sequential ()  
Model ← Model.add(Embedding(Wc, Ed, Lm, e))  
Model ← Model.add(Conv1D(64, 3, Ac = 'sigmoid'))  
Model ← Model.add(CNN(128))  
Model ← Model.add(BiLSTM(128))  
Model ← Model.add(BiGRU(128))  
Model ← Model.add(Flatten())  
Model ← Model.add(Dense(1, Ac = '77'))
```

Compilation of Model:

```
Model.compile('binary_crossentropy', 'Adam', 'Accuracy')
```

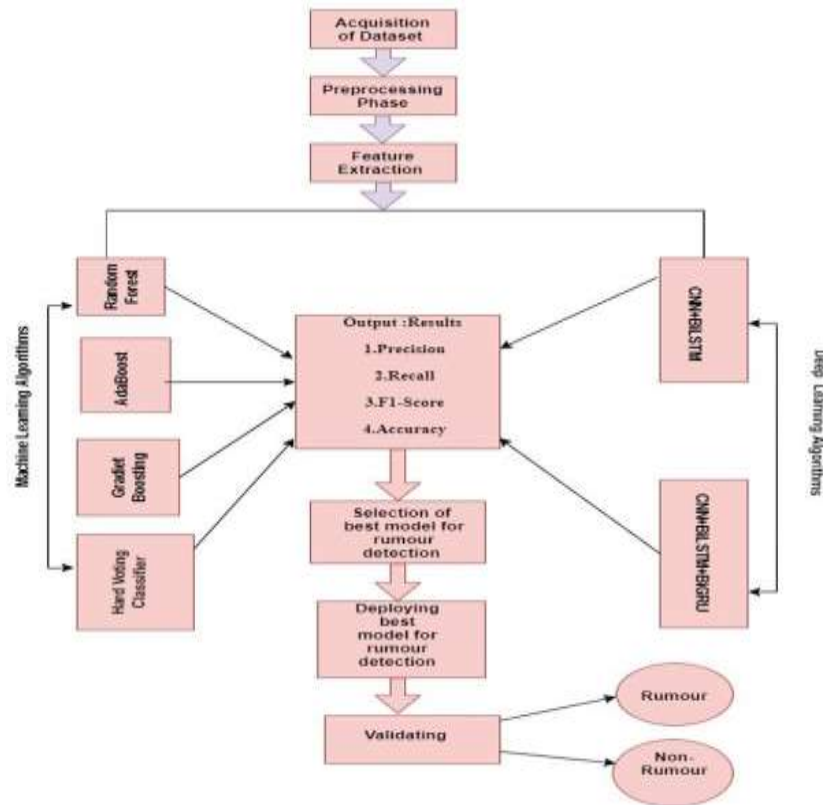
Model Training & Evaluation:

```
Hist ← Model.fit (Tp, 126, 60, 0.40)
```

Numerous deep learning techniques are integrated into the suggested hybrid model. We applied deep learning algorithms (CNN, BiLSTM, and BiGRU) to the dataset and assessed their results following the first steps of dataset extraction, preprocessing, feature extraction, and data tokenisation. Our goal is to determine which model performs better.

CNN examines images and videos. A multilayer perceptron is equivalent to a convolutional neural network (CNN) [20]. Every neurone in the first layer of a multilayer perceptron is connected to every neurone in the subsequent layer. Because CNN's neurones learn on their own, relatively little pre-processing is required. CNN performs well in a

number of tasks, including text and picture sorting. Since CNN is quicker and more efficient than earlier deep learning techniques, researchers have recently attempted to determine how CNN influences NLP [21–22]. The hidden state (ht) of a unidirectional LSTM only takes into account past data; it makes no reference to future stages. In a BiLSTM design, forward and backward LSTM layers are considered effective at extracting information from previous and subsequent states.



IV.Results and Discussions

These metrics are referred to as performance evaluation metrics. The accuracy of the suggested hybrid model is the primary metric used in this study to assess its effectiveness. To verify the efficacy of the suggested methodology, the metrics derived from the confusion matrix will be contrasted with other classification performance criteria. Precision, recall, F1-score, and accuracy are all provided by the confusion matrix. The confusion matrix for the classification problem is displayed in Figure 4. "Positive" in the confusion matrix denotes accurate knowledge, while "negative" refers to rumors. The following definitions apply to the acronyms shown in the confusion matrix.

True-Positive (TP): TP denotes that the actual and predicted classes are both positive.



True-Negative (TN): The actual numbers are negative while the expected class outcome is positive.

False-Positive (FP): The actual numbers are negative even though the projected class output is negative.

False-Negative (FN): When real values are positive but the anticipated class value is negative.

Accuracy: It is the most often used and possibly the most important metric for evaluating the performance of algorithms in classification problems. The following equation is used to forecast accuracy. The accuracy indicates the percentage of accurate forecasts.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Precision: It shows how many of the chosen data points are relevant. It is the proportion of relevant instances to instances that were retrieved. Precision is defined as the ratio of correct predictions of false information to all rumors.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

Recall: It shows how many relevant data points are chosen. It is the percentage of relevant cases that are successfully retrieved. It represents the proportion of correct results to all retrieved results. The recall represents the percentage of false information found in relation to all anticipated rumors.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

F1-Score: The F1-Score, sometimes called an f-score or an f-measure, is a measure of an algorithm's performance that depends on its recall and precision respectively. The F1-score is another metric that can be used to track accuracy. Both recall and precision are incorporated into it.

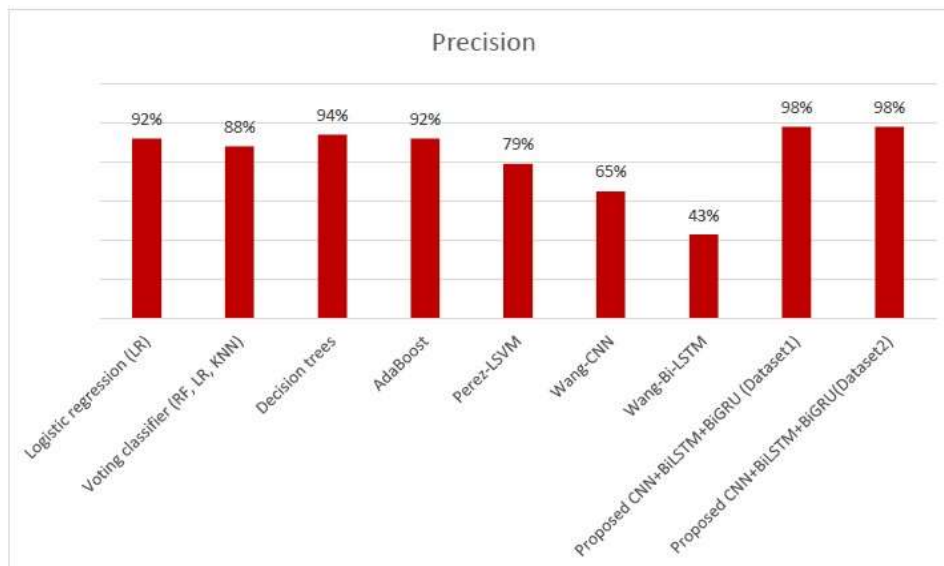
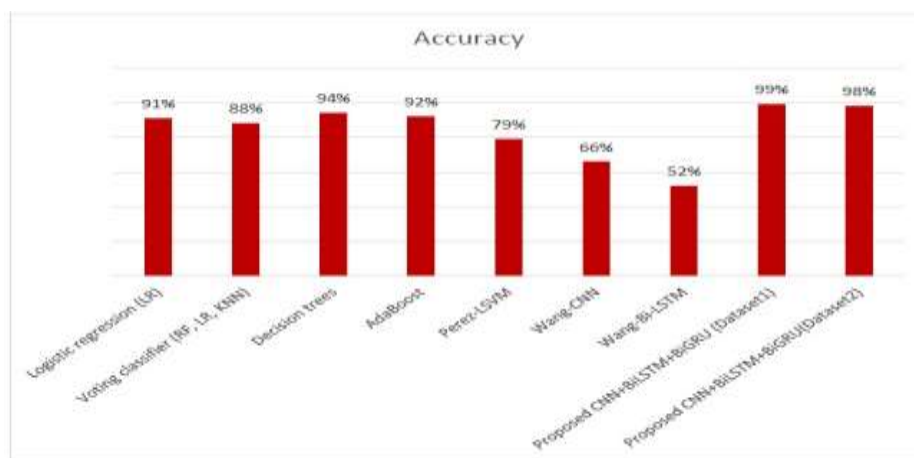
$$\text{F1-Score} = 2 * ((\text{precision} * \text{recall}) / (\text{precision} + \text{recall})) \quad (4)$$

The model's performance in a particular scenario can be evaluated using the accuracy, precision, recall, and F1-score performance metrics mentioned above. When it comes to making predictions in novel circumstances, these help us identify what these models excel at and what they struggle with. For both machine learning and deep learning, model performance is crucial.

		Actual Value	
		Positive	Negative
Predicted Value	Positive	TP	FP
	Negative	FN	TN

CNN+BiLSTM+BiGRU

This section of the article demonstrates how the dataset is used by the suggested hybrid model. Additionally, we contrast our model with the previous study's model using the same dataset. The confusion matrix for the proposed framework is displayed in Figure 4, and performance Table-II compares accuracy, precision, recall, and f1-score. The same pattern can be seen in the comparison based on accuracy, precision, and recall in Figure 5. The performance of the suggested method is significantly superior to that of alternative algorithms. AdaBoost is a strong competitor to our proposed approach. With a score of about 99, the recommended algorithm performs well across the board. Decision Trees, the closest competitor, has a score of about 95, which is significantly lower than that of the suggested algorithm



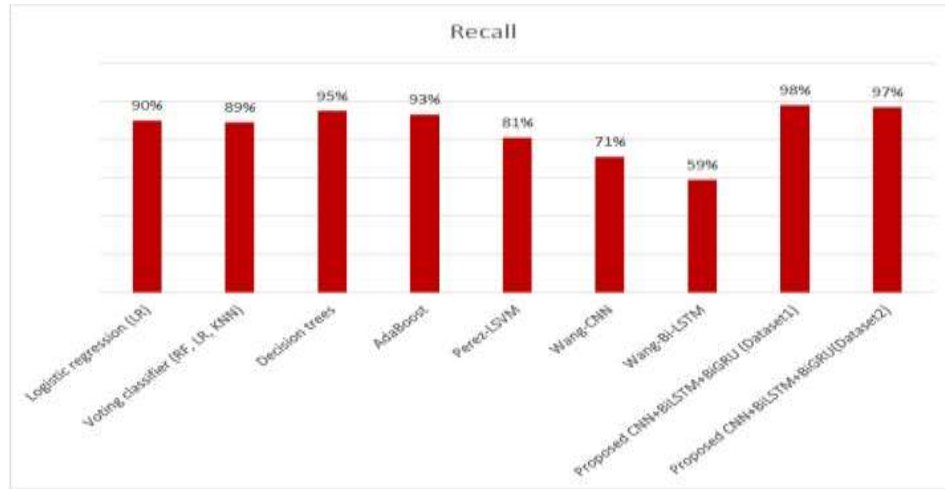


Fig. 5. Comparisons With Previous Works

Table Ii : Performance Index

Methods	Accuracy	Precision	Recall	F1-score
Logistic-Regression	92 %	93 %	92 %	93 %
Voting-Classifier	90 %	89 %	90 %	89 %
Decision-Trees	93 %	95 %	94%	93 %
AdaBoost	95%	94%	94%	95 %
Perez-LSVM	81%	82%	80%	82%
Wang-CNN	68%	69%	73%	71%
Wang-Bi-LSTM (Dataset)	69%	67%	62%	65%
Proposed CNN+BiLSTMs+BiGR U	98.5%	97%	96%	99%

V. Conclusion And Future Works

This research covers a variety of sophisticated machine learning and deep learning approaches for addressing the problem of rumour detection. A sophisticated hybrid deep learning model was created to construct a strategic link inside the text data. The suggested method employs a unique strategy that combines bespoke embedding with the CNN, BiLSTM, and BiGRU algorithms. Various attributes were fed into the sigmoid activation function for categorisation. The dataset was used to conduct an experimental assessment of the proposed model, followed by a performance comparison with other



existing models. The proposed hybrid model achieved an accuracy of 98.5%; the greatest result compared to previous research.

The results show that word embedding techniques increased the proposed method's (CNN+BiLSTM+BiGRU) classification accuracy in distinguishing tweets as rumours or non-rumors. The proposed task has various disadvantages; yet, it also has some advantages.

- I. The rumour classification in this investigation was entirely based on textual elements. The use of additional parameters may produce more robust results.
- II. In addition to text-based qualities, images and contextual information may improve outcomes.
- III. Future examinations of the textual content will take into account linguistic views. We plan to investigate several deep-learning methods for spotting rumours..

References

- Barsha Pattanaik, et al. 2023. A survey on rumor detection and prevention in social media using deep learning: Survey on rumor detection and prevention. *Knowl. Inf. Syst.* 65, 10, 3839–3880. <https://doi.org/10.1007/s10115-023-01902-w>
1. Mohammad Ziari, et al. 2026, Rumor detection and propagation on social networks: A survey, *Expert Systems with Applications*, Volume 295, in *Expert Systems with Applications*, Elsevier, <https://doi.org/10.1016/j.eswa.2025.128798>.
 2. Zhang, L., Chen, Y., Luo, X. et al. 2026 Early detection of rumors based on propagation prediction in social media. *Complex Intell. Syst.* 12, 23. <https://doi.org/10.1007/s40747-025-02140-z>
 3. Ayyasamy RK, et al. 2021. A hybrid deep learning framework for fake news detection using LSTM-CGPNN and metaheuristic optimization. *Sci Rep.* <https://doi.org/10.1038/s41598-025-25311-x>.
 4. M. F. MRIDHA et al. 2021, A Comprehensive Review on Fake News Detection With Deep Learning, *IEEE Access*, <https://doi.org/10.1109/ACCESS.2021.3129329>.
 5. Jabir, H.M., Naser, et al. 2020. Rumour detection on twitter using features extraction method. In. *Information Technology to Enhance e-learning and Other Application (IT-ELA)* (pp. 115-120). IEEE.
 6. Vijeev, A., Mahapatra, A., Shyamkrishna, A. and Murthy, S., 2018, September. A hybrid approach to rumour detection in microblogging platforms. In *2018 International Conference on Advances in Computing, Communications and Informatics* (pp. 337-342). IEEE.
 7. Cheong, F. and Cheong, C., 2011, Social Media Data Mining: A Social Network Analysis Of Tweets During The 2010-2011 Australian Floods. *PACIS*, 11, pp.46-46.



8. Nguyen, T.N., Li, C. and Niederée, C., 2017, September. On early-stage debunking rumours on twitter: Leveraging the wisdom of weak learners. In International Conference on Social Informatics (pp. 141-158). Springer, <https://doi.org/10.48550/arXiv.1709.04402>
9. Ma, J., Gao, W. and Wong, K.F., 2018. Rumour detection on twitter with tree-structured recursive neural networks. Association for Computational Linguistics.
10. Sadiq, S., Mehmood, A., Ullah, S., Ahmad, M., Choi, G.S. and On, B.W., 2021. Aggression detection through deep neural model on twitter. Future Generation Computer Systems, 114, pp.120-129, <https://doi.org/10.1016/j.future.2020.07.050>
11. Ajao, O., Bhowmik, D. and Zargari, S., 2018, July. Fake news identification on twitter with hybrid cnn and rnn models. In Proceedings of the 9th international conference on social media and society (pp. 226-230).
12. Gao, J., Han, S., Song, X. and Ciravegna, F., 2020. Rp-dnn: A tweet level propagation context based deep neural networks for early rumour detection in social media. arXiv preprint arXiv:2002.12683.
13. Alkhodair, S.A., Ding, S.H., Fung, B.C. and Liu, J., 2020. Detecting breaking news rumours of emerging topics in social media. Information Processing & Management, 57(2), p.102018, <https://doi.org/10.1016/j.ipm.2019.02.016>.
14. Asghar, M.Z., Habib, A., Habib, A., Khan, A., Ali, R. and Khattak, A., 2021. Exploring deep neural networks for rumour detection. Journal of Ambient Intelligence and Humanized Computing, 12(4), pp.4315-4333, <https://doi.org/10.1007/s12652-019-01527-4>
15. Chen, T., Xu, R., He, Y. and Wang, X., 2017. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. Expert Systems with Applications, 72, pp.221- 230.
16. Zeng, H., Wang, R., Huang, Y., Cui, X. and Jiang, Q., 2021, October. Scientific Rumours Detection in Short Online Texts. In 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 1233-1240), <https://doi.org/10.1109/SMC52423.2021.9659056>.
17. Sridhar, S. and Sanagavarapu, S., 2021, January. Fake news detection and analysis using multitask learning with BiLSTM CapsNet model. In 11th International Conference on Cloud Computing, Data Science & Engineering (pp. 905-911). IEEE.
18. Rehman, A.U., Malik, A.K., Raza, B. and Ali, W., 2019. A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis. Multimedia Tools and Applications, 78(18), pp.26597-26613, <https://doi.org/10.1007/s11042-019-07788-7>.
19. Khan, S., Kamal, A., Fazil, M., Alshara, M.A., Sejwal, V.K., Alotaibi, R.M., Baig, A.R. and Alqahtani, S., 2022. HCovBi-caps: hate speech detection using convolutional and Bi- directional gated recurrent unit with Capsule network. IEEE Access, , pp.7881-7894, <https://doi.org/10.1109/ACCESS.2022.3143799>.
20. Fazil, M., Sah, A.K. and Abulaish, M., 2021. Deepsbdb: a deep neural network model with attention mechanism for socialbot detection. IEEE Trans. on Information Forensics and Security, 16, pp.4211-4223, <https://doi.org/10.1109/TIFS.2021.3102498>.



21. Salur, M.U. and Aydin, I., 2020. A novel hybrid deep learning model for sentiment classification. IEEE Access, 8, pp.58080-58093, <https://doi.org/10.1109/ACCESS.2020.2982538>.
22. Sandeep Sahu, Ravi Verma, Vijay Kumar Trivedi (2025). "A Review on IoT-enabled Body Sensor Networks for Healthcare System", in the book Applications of Deep Learning in Genomics (1st ed.). CRC Press. <https://doi.org/10.1201/9781003558835>.