Journal of Ambient Intelligence and Humanized Computing (2023) 14:4451–4460
https://doi.org/10.1007/s12652-023-04562-4

ORIGINAL RESEARCH

A cooperative deep learning model for fake news detection in online social networks

Chandrakant Mallick[1] · Sarojananda Mishra[2] · Manas Ranjan Senapati[3]

Received: 21 February 2022 / Accepted: 8 February 2023 / Published online: 28 February 2023

Abstract
Fake news, which considers and modifies facts for virality objectives, causes a lot of havoc on social media. It spreads faster than real news and produces a slew of issues, including disinformation, misunderstanding, and misdirection in the minds of readers. To combat the spread of fake news, detection algorithms are used, which examine news articles through temporal language processing. The lack of human engagement during fake news detection is the main problem with these systems. To address this problem, this paper presents a cooperative deep learning-based fake news detection model.The suggested technique uses user feedbacks to estimate news trust levels, and news ranking is determined based on these values. Lower-ranked news is preserved for language processing to ensure its validity, while higher-ranked content is recognized as genuine news. A convolutional neural network (CNN) is utilized to turn user feedback into rankings in the deep learning layer. Negatively rated news is sent back into the system to train the CNN model. The suggested model is found to have a 98% accuracy rate for detecting fake news, which is greater than most existing language processing based models.The suggested deep learning cooperative model is also compared to state-of-the-art methods in terms of precision, recall, F-measure, and area under the curve (AUC). Based on this analysis, the suggested model is found to be highly efficient.

Keywords Social media · Fake news · Deep learning · Cooperative · Convolutional · Language processing

1 Introduction

People can now quickly obtain news through multiple internet-based platforms such as social media, blogs, and webpages.These sources include a huge amount of unverified and unauthenticated content, leading to widespread misinformation. Fake news is an example of such unconfirmed and unauthenticated information. Fake news are prepared or released with the goal to deceive the public and harm the image of a company, organisation, or individual, either for financial or political gain (Kaliyar et al. 2021a, b). Due to the abundance of various online platforms, the circulation of fake news has become easier. Individuals using these platforms can make and distribute fake news content depending on personal or professional motif. The concept of fake news in online social networks has emerged like wildfire due to its chaotic effect on human behavior. Fake news can be any form of misleading material, fake reviews, fake rumours, ads, fake political speech, satires, and other forms (Sahoo and Gupta 2021). For instance, there were numerous news articles regarding the recent outbreak of CoVID-19 that has affected the entire world, wherein fake news were spread purposely through twitter and facebook. Fake news are widely distributed on social media sites, misinforming the public, disturbing social order, undermining government confidence, and posing a serious threat to societal stability. Stock markets can be forecasted using data from social media and financial news (Khan et al. 2020). Sentiment analysis, also known as opinion mining using social media data,

✉ Chandrakant Mallick
ckmallick@gmail.com

Sarojananda Mishra
sarose.mishra@gmail.com

Manas Ranjan Senapati
manassena@gmail.com

1 Department of Computer Science and Engineering, Biju Patnaik University of Technology, Rourkela, Odisha 769015, India

2 Department of Computer Science and Engineering, Indira Gandhi Institute of Technology, Sarang, Odisha 759146, India

3 Department of Information Technology, Veer Surendra Sai University of Technology, Burla, Odisha 768018, India

is a technique for mining the public's opinions or feelings in order to forecast election outcomes (Chauhan et al. 2021). So fake news is capable of directing election results, causing riots, destabilizing stocks, and possesses various other disastrous capabilities.The propagation of fake news has a very harmful influence on the targeted individuals and the society as a whole. As a result, it produces a sense among readers that the overall perception and response to legitimate news is diluted and trustworthiness of the news channels is degraded (Hakak et al. 2021).

In order to improve the trustworthiness of the online platforms such as online social networks and reduce the disastrous impacts on the society, it is essential to develop a reliable mechanism for early detection and containment of misleading contents like fake news. Many endeavours research based on supervised learning methods were made for the issue of detecting fake news. Many of those works suffer from some limitations of low accuracy. The basis for low accuracy can be caused by many reasons such as the mediocre feature selection, incompetent parameter tuning, non-availability of benchmarked and balanced datasets, etc. Some of the researchers propose comparatively accurate fake news detection and classification models using deep learning techniques to reduce the probability of these issues. One of the main reason of using deep learning techniques is that in traditional machine learning techniques,human intervention is used explicitly to do feature engineering. However feature engineering is not needed in deep learning because important features are automatically detected by deep neural networks in deep learning. Among the deep learning techniques, the convolutional neural network (CNN) is more popular in recent research because it automatically detects the important features without any human supervision. It has been evidenced in the literature that in many cases CNN models outperform than other contemporary models.

A general-purpose architecture followed by our fake news detection model using language processing and deep learning can be observed from Fig. 1. Based on the general-purpose architecture of the model, a wide variety of machine learning and deep learning methods using language processing are proposed by researchers for fake news detection. A survey of these deep learning models is presented in the next section.

In this study we propose a deep learning based cooperative model that uses the user feedbacks to estimate news trust levels, and news ranking is determined based on these values of the trust levels. Lower-ranked news is preserved for language processing to ensure its validity, while higher-ranked content is recognised as genuine news. A deep learning model which is a variant of convolutional neural network (CNN) known as VGG 16 (Visual Geometry Group with 16 deep network layers) is utilized to turn user feedbacks into rankings in the deep learning layer. Negatively rated news is sent back into this standard CNN model to train it. The experimental results shows that the cooperative model outperformed in comparison to the state of the art.

The model starts with collecting news articles from the available online datasets which are prepared by scraping

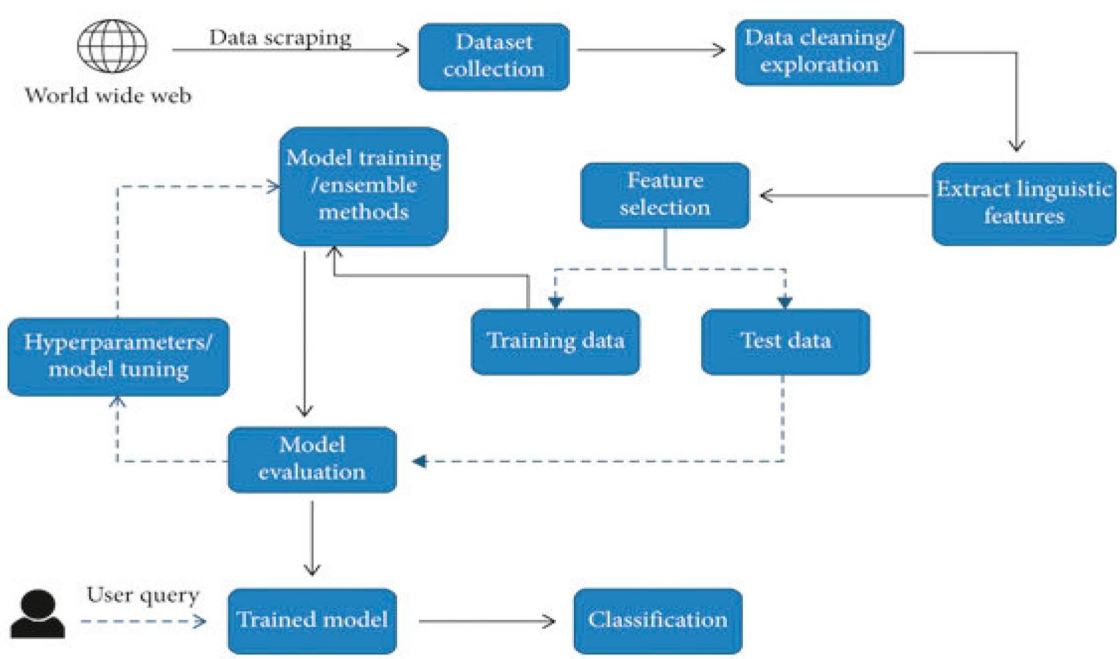

Fig. 1 General-purpose architecture of fake news detection model (Ahmad et al. 2020)

news data from the internet (or world wide web) and prepares a dataset of these collected news articles.

The articles in the prepared dataset are processed using the following steps.

- *Data cleaning and exploration* which is used to remove any repeated or outlier news articles from the collected dataset.
- *Linguistic feature extraction* Applies word embedding techniques like “word2vec” and 'n' gram approaches which are used to convert the news articles into feature vectors.
- *Feature selection* A feature selection model is used that removes non-variant features to reduce the time delay in fake news detection process and improve the accuracy of classification models. We have used the method of independent component analysis (ICA) for the purpose of feature selection.
- *Train and test data* The dataset with selected features are divided into training and testing data, out of which training data is used for linguistic learning.
- *Model training and evaluation* In this step, learning is performed using ensemble models, or deep learning models wherein the model evaluation is performed. Here, we used the standard CNN model VGG 16.
- *Model tuning* The result of these models is used for accuracy evaluation and hyperparameter tuning via feedback-based learning.
- *User Query* A tuned model is used to check user input queries and estimate if the input query results in fake news.
- *Classification* The results obtained from this classifier are given to a feedback model, wherein confidence-based checking is done and the model incrementally learns from its classified output.

The performance of the proposed cooperative deep learning model for detection of fake news using user feedback is now evaluated and compared with the state of the art.

Rest of this paper is organized as follows. Section 2 includes a review of the related works in the literature, Sect. 3 explains the design of the proposed cooperative deep learning based fake news detection model, Sect. 4 includes the description of performance evaluation of the proposed model and its comparison with various state-of-the-art methods, Sect. 5 presents the results interpretation. Finally, Sect. 6 concludes with some interesting observations about the proposed model and recommends methods to improve its performance in future.

2 Literature review

Researchers have developed a wide variety of deep learning-based fake news detection systems. For instance, the work by Ni et al. (2021), Ali et al. (2021) and Verma et al. (2021) validate the use of multiple views attention networks (MVNN), adversarial networks, and word embedding feature extraction for fake news identification respectively. These models can achieve an accuracy of over 80% for fake news detection, which makes them deployable for moderately sized systems. To improve this performance, the work by Han et al. (2021) proposes the use of a two-stream network for fake news detection. Similarly the work by Li et al. (2021) uses unsupervised fake news detection method based on auto encoder, and the work by Jiang et al. (2021) proposes an ensemble method of stacking of logistic regression, decision tree, k-nearest neighbor, random forest, and support vector machine (SVM). All of these approaches achieved an accuracy of over 85%, for verification of real-time generated news. These models find their applications in various systems, including private sector news, public sector news, and government news broadcast channels. The speed of detection is slow in these network models due to high complexity neural networks. This speed can be improved using the work by Dong et al. (2020), Sansonetti et al. (2020) and Elhadad et al. (2020) wherein two-path deep semi-supervised learning, a deep neural network for fake user detection, and decision tree models are deployed. These models have moderate accuracy performance but are highly effective for quick decisions during real-time fake news detection scenarios.

The efficiency of fake news detection is improved using the work by Umer et al. (2020) and Shrivastava et al. (2020) wherein CNN is combined with long short term memory (LSTM) and defensive modeling via online social networks (OSNs). These models are able to achieve an accuracy of over 89% for fake news detection under certain categories. But the average accuracy varies between 59 and 85% for various out-of-trend categories like corridor news, parking news, etc. This performance can be improved via the use of category-specific fake news detection models. Such models are demonstrated by Subbanarasimha et al. (2020), De Oliveira et al. (2020) and Neves et al. (2020), wherein social cognition, social and blogging networks, and multimedia fake news detection methods are described. These methods can identify category-specific fake news with over 90% accuracy, making them suitable for internal component design when developing a large-scale fake news detection system. Similar models are proposed by Dutta and Chakraborty (2020) and Xu et al.

(2020), wherein darknet and social media platforms are analyzed for fake news using deep learning methods. Similarly, Raj and Meel (2021) suggested a multi-modal fake news detection model that included Text-CNN and Image-CNN modules on two separate text datasets, wherein their Text-CNN module performs averaged 94.91% accuracy. The authors Shim et al. (2021) suggested using the composition structure of web links containing news content as a new source of information for detecting fake news. They suggested a unique embedding technique called link2vec, which is an extension of word2vec, to properly vectorize the composition pattern of web links and increase classification accuracy. Further they used ensemble machine technique to classify fake news. They achieved the maximum accuracy of 93.1%. Authors Mehta et al. (2021), proposed a model for fake news identification based on natural language processing and Bidirectional Encoder Representations from Transformers (BERT), which introduced the multiple BERT layer neural networks and obtained 74% accuracy. Table 1 represents the summary of the detection approaches, performances and relative limitations of the related works on fake news detection.

Based on this analysis, it is observed that deep learning CNN Models outperform linear fake news detection models. So we proposed a model that is designed using a cooperative feedback mechanism and the deep learning model called VGG 16 CNN for fake news detection. The description of the proposed model can be observed from the next section.

3 Proposed methodology

Observations from the literature review indicate that deep learning models outperform linear language processing models in terms of the accuracy of fake news detection. Thus, this text proposes the design of a cooperative deep learning model that uses VGG 16 CNN model to estimate news rank. There are many CNN-based fine tuned models such as VGG16, VGG19, ResNet, ShallowNet, DenseNet, XceptionNet, NASNet**.** However, VGG 16 model has been applied effectively for image classification and has been proved to be the best performing model on ImageNet dataset. We adopted the VGG 16 model as it can work better for language processing tasks like fake news detection.

The proposed methodology uses the following steps to decide whether a given news article is genuine or fake.

(1) For every new article posted or crawled, evaluate the following parameters,

- o Location of news (N_{loc})
- p Category of news ($N_{category}$)
- q Sentiment of news ($N_{sentiment}$)

Table 1 Summary of the performances of the related fake news detection models

References	Datasets	Detection approaches	Accuracy (%)	Limitations
Henrique et al. (2020)	FakeBrCorpus, FakeOrReal-News	Convolutional Neural Network (CNN)	79.00	Higher accuracy is not attained
Asghar et al. (2021)	COAE2014, IMDB	BiLSTM-CNN + task-specific word embedding	86.12	Used text-based features on English language only
Aslam et al. (2021)	LIAR dataset	Bi-LSTM-GRU-dense deep learning model	89.8	Used only one dataset.The model needs to be validated using other datasets
Kumar et al. (2019)	PolitiFact	CNN + Bi- LSTM ensembled	88.78	Other recently developed architectures can provide better accuracy
Kaliyar et al. (2021a, b)	PHEME	C-LSTM – A Hybrid model of CNN and LSTM	91.88	The model does not achieve similar accuracy in different datasets
Thota et al. (2018)	FNC-1	Dense neural network (DNN)	94.31	To be verified on different real-time social network datasets
Varshney and Vishwakarma (2021)	LIAR	Hybrid feature extraction, Random Forest classifier	95.00	Only one dataset is taken. It could use any multimedia data
Liu and Wu (2020)	Weibo and Twitter	Convolutional Neural Network (CNN)	90.00	A small dataset of 1,111 Twitter posts and 816 Weibo posts was used
Li et al. (2019)	LIAR, Weibo, Twitter15 and KaggleFN	Multiple-level convolutional neural network (MCNN)	91.87	Data collected related to cultural communication only. It could be applied on wider range of other applications

r Entities (nouns and pronouns) mentioned in the article ($N_{entities}$)

(2) Broadcast this news to every user who is in the vicinity of N_{loc} using the following Eq. 1.

$$H_{d_i} = R * 2 * \left(\tan\left(\sin^2\left(\frac{lat_i - lat_{article}}{2}\right) + \sin^2\left(\frac{lon_i - lon_{article}}{2}\right) * \cos\left(lat_i\right) * \cos\left(lon_{article}\right)\right)\right)^{1/2} \ldots(1) \quad (1)$$

where, $H_{d_i}, R, lat, andlon$, represents Haversine distance between N_{loc} and user, radius of Earth, latitude, and longitude of the nodes.

(C) Verify feedback about this news article from each user, and aggregate their responses using Eq. 2,

$$AGG_{responses} = \frac{\sum_{i=1}^{N_n} F_{news}}{N_n} ...(2) \quad (2)$$

where, $N_n, AGG_{responses}, andF_{news}$ represents the number of news articles in the vicinity, aggregate responses, and feedback about the genuineness of the current news article.

(D) Evaluate the temporal feedback about each identified entity in the given category ($N_{category}$), and correlate it with a sentiment of the current news article.
(E) Sentiment analysis is done using the text blob method, wherein each word and its sentiment is aggregated to estimate the sentence's final sentiment.
(F) Evaluate this correlation using Eq. 3.

$$C = \frac{\sum_{i=1}^{N_{entities}} S_{i_{news}} - S_{i_{overall}}}{\sqrt{\sum_{i=1}^{N_{entities}} \left(S_{i_{news}} - S_{i_{overall}}\right)^2}} \quad (3)$$

where, $S_{i_{news}}, andS_{i_{overall}}$ represents the sentiment of the current entity in the news article and the sentiment of the existing entity in all the fetched news articles to date.

(G) Provide all these features, Location of news (N_loc), Category of news (N_category), Sentiment of news (N_sentiment), Entities (nouns and pronouns) mentioned in the article (N_entities) as input to the standard VGG16 CNN model, and estimate its ranking.
(H) The VGGNet16 model is initially trained with the extracted features and approximate ranking of the news article. The layered architecture of the adapted VGG16CNN model can be observed in Fig. 2.
(I) The model uses a combination of convolutional and pooling layers to augment news features and estimate the ranking of the news.
(J) All correctly classified test set entries are given feedback into the system to retrain the network for better accuracy.

Based on this network design, the system is evaluated for different types of news articles. The classification performance evaluation and comparison with standard models can be observed from the next section.

4 Performance evaluation

4.1 Dataset

The dataset is created by collecting articles from three datasets that are available online and have been extracted directly from the World Wide Web. The first dataset is the ISOT Fake News dataset, followed by the "Fake News" and "Fake News Detection" datasets, all of which are freely accessible on Kaggle.com. The dataset created is a mix of real and fake news pieces and include news from several different domains such as politics, sports, finance,

Fig. 2 Layered architecture of VGG16 CNN Model

technology etc.The new dataset is built to evaluate the performance of algorithms on datasets which include a wide range of fields in a single dataset because the articles differ in type in each dataset. Nearly 100 recent articles were extracted from the existing datasets for each category for this evaluation. About 70% of the news articles from the created dataset were used for training and 30% of the articles were used for testing purposes.

4.2 Performance metrics

A large number of news articles were given as input for fake news detection to the proposed model, and upon retrieval of the results of classification in terms of the confusion matrix, the values of the performance metrics, accuracy (A), precision (P), recall (R), and F-measure (F) were evaluated using the following equations.

$$A = \frac{T_p + T_n}{T_p + F_p + T_n + F_n} \dots (4) \tag{4}$$

$$P = \frac{T_p}{T_p + F_p} \dots (5) \tag{5}$$

$$R = \frac{T_p}{T_p + F_n} \dots (6) \tag{6}$$

$$F = \frac{2 * P * R}{P + R} \dots (7) \tag{7}$$

where, T_p is true positive value, which is the number of news articles correctly classified as positive for a particular news category,T_n is the true negative value which is the number of news correctly classified as negative for a particular news category,F_p is false positive value, which is the number of news articles that belong to the given category but is incorrectly classified, and F_n is false-negative value, which is the number of articles that do not belong to the given category, and are incorrectly classified. Different query articles were given to the system based on these values, and the A, P, R, and F values were observed. This observation is tabulated for different article types in Table 2, wherein average values for accuracy are evaluated for the models, Naive Bayesian (NB) and Support Vector Machines (SVM), and the proposed model. The graphical representation of accuracy values of different news categories for three different models is presented in Fig. 3.

It can be observed that the proposed model is nearly 12% better in terms of average accuracy values when compared with NB and around 10% better when compared with SVM for the given categories. Similarly, values for precision were evaluated and tabulated using Table 3 as follows. The graphical representation of Precision values of different news categories is presented in Fig. 4.

Table 2 Accuracy values for different news category

News category	Avg. A [NB]	Avg. A [SVM]	Avg. A [Proposed]
Agriculture	0.91	0.95	0.95
Aviation	0.53	0.61	0.68
Sports	0.59	0.65	0.72
Leisure	0.42	0.46	0.46
Roads	0.68	0.72	0.80
Construction	0.63	0.67	0.72
Residential	0.57	0.61	0.61
Forest	0.61	0.61	0.65
Highways	0.68	0.68	0.72
Outdoor sports	0.72	0.72	0.76
International Waters	0.57	0.61	0.61
City Roads	0.53	0.57	0.57
Village	0.46	0.53	0.57
Park news	0.34	0.42	0.46
Corridor news	0.87	0.91	0.91
Parking news	0.91	0.95	0.95
Finance	0.91	0.95	0.95
Politics	0.91	0.91	0.91
Local	0.86	0.87	0.91
Technology	0.46	0.53	0.53
Indoor sports	0.68	0.72	0.76

It can be observed that the proposed model is nearly 8% better in terms of average precision values when compared with NB model and around 14% better when compared with SVM for the given categories. Similarly, the values of recall were evaluated and tabulated in Table 4 as follows. The graphical representation of Recall values of different news categories is presented in Fig. 5.

It can be observed that the proposed model is nearly 14% better in terms of average recall values when compared with NB model and around 8% better when compared with SVM model for the given categories. Similarly, the values of the F-measure were evaluated and tabulated in Table 5 and the graphical representation of F-Measure values of different news categories is presented in Fig. 6.

It can be observed that the proposed model is nearly 15% better in terms of average F-measure values when compared with NB model and around 6.5% better when compared with SVM model for the given categories. Similarly, the values of AUC were evaluated and tabulated in Table 6 and graphical

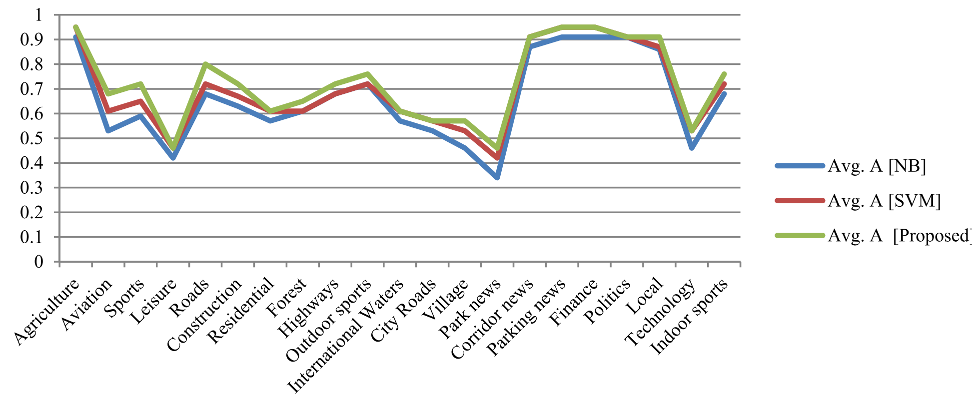

Fig. 3 Graphical representation of accuracy values of different news categories

Table 3 Precision of different news categories

News category	Avg. P [NB]	Avg. P [SVM]	Avg. P [Proposed]
Agriculture	0.94	0.98	0.98
Aviation	0.55	0.63	0.71
Sports	0.61	0.67	0.74
Leisure	0.43	0.47	0.47
Roads	0.71	0.74	0.82
Construction	0.65	0.69	0.74
Residential	0.59	0.63	0.63
Forest	0.63	0.63	0.67
Highways	0.71	0.71	0.74
Outdoor sports	0.74	0.74	0.78
International Waters	0.59	0.63	0.63
City Roads	0.55	0.59	0.59
Village	0.47	0.55	0.59
Park news	0.35	0.43	0.47
Corridor news	0.90	0.94	0.94
Parking news	0.94	0.98	0.98
Finance	0.94	0.98	0.98
Politics	0.94	0.94	0.94
Local	0.88	0.90	0.94
Technology	0.47	0.55	0.55
Indoor sports	0.71	0.74	0.78

representation of AUC values for different news categories is given in Fig. 7.

It can be observed that the proposed model is nearly 15% better in terms of average AUC values when compared with NB model and around 8% better when compared with SVM model for the given categories.

5 Results interpretation

The findings for the proposed model for classification of fake and real news can be observed in the tables and figures in the previous section. In this research, we employed a cooperative deep learning-based model with VGG 16 CNN and evaluated its performance using a variety of metrics, such as accuracy, precision, recall, and the F-measures. According to the experimental results, the proposed model scored the maximum accuracy of 98%. The results show that the proposed model is nearly 12% better in terms of average accuracy values when compared with NB model and around 10% better when compared with SVM model for the given news categories. Also, the proposed model is nearly 14% better in terms of average precision values when compared with NB and around 8% better when compared SVM model for the given news categories. For the specified categories, it can be seen that the suggested model performs around

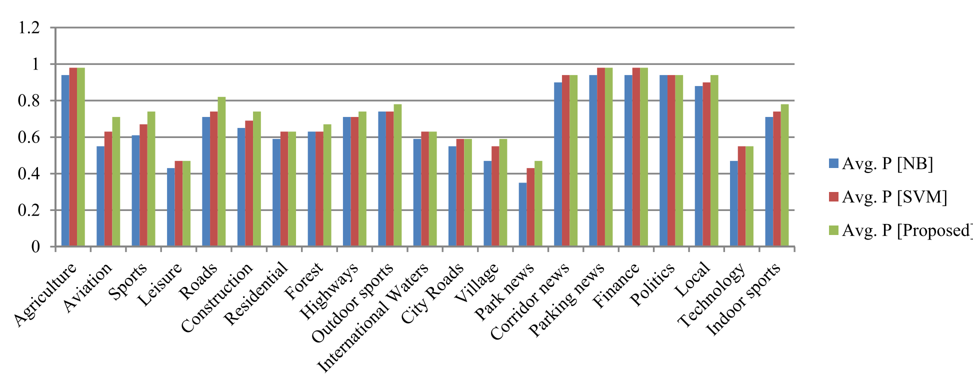

Fig. 4 Graphical representation of precision values of different news categories

Table 4 Recall values for different news categories

News category	Avg. R [NB]	Avg. R [SVM]	Avg. R [Proposed]
Agriculture	0.47	0.51	0.61
Aviation	0.27	0.35	0.53
Sports	0.55	0.61	0.78
Leisure	0.18	0.22	0.31
Roads	0.14	0.18	0.35
Construction	0.52	0.56	0.72
Residential	0.29	0.33	0.43
Forest	0.57	0.57	0.71
Highways	0.49	0.49	0.63
Outdoor sports	0.74	0.74	0.88
International Waters	0.59	0.63	0.73
City Roads	0.47	0.51	0.63
Village	0.37	0.45	0.59
Park news	0.25	0.32	0.46
Corridor news	0.73	0.76	0.86
Parking news	0.61	0.65	0.73
Finance	0.57	0.61	0.71
Politics	0.57	0.57	0.67
Local	0.53	0.55	0.69
Technology	0.42	0.50	0.60
Indoor sports	0.71	0.74	0.88

8% better than the SVM model and about 15% better than the NB model in terms of average AUC values. The results clearly demonstrate that our suggested model is excellent at classifying fake news on news articles in different domains.

6 Conclusion and future work

In our research, we have experimented using two machine-learning algorithms such as NB and SVM and one cooperative deep learning model with VGG 16 CNN to detect fake news on the dataset prepared from three popular datasets available online which contains news from different domains. The performance of the classifier models has been analyzed in the aspects of accuracy, precision, recall, and F-measures for all the algorithms. The AUC values of different models are also compared for each category of news articles. Based on this performance, it can be concluded that our suggested cooperative deep learning model is better than the traditional machine learning techniques and our model is helpful for high accuracy real-time fake news classification. The model finds itself to be useful for a wide variety of scenarios, including the implementation of Open Journalism, wherein every device can be used as a media gathering interface with high trust levels. In the future, it is recommended that the proposed model must be implemented with high speed and low complexity of news verification. In our

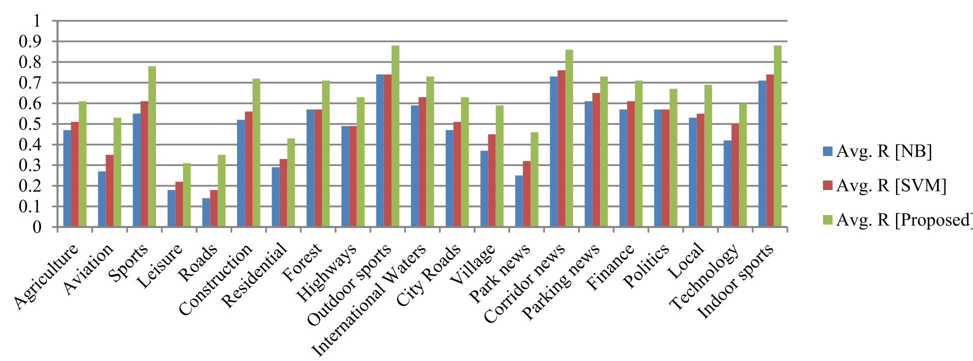

Fig. 5 Representation of Recall values for different news categories

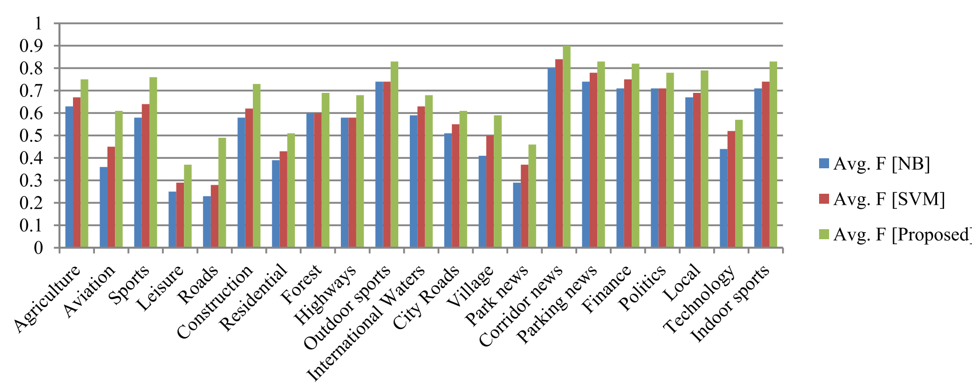

Fig. 6 Representation of F-measures values for different news categories

Table 5 F-Measure values for different news categories

News category	Avg. F [NB]	Avg. F [SVM]	Avg. F [Proposed]
Agriculture	0.63	0.67	0.75
Aviation	0.36	0.45	0.61
Sports	0.58	0.64	0.76
Leisure	0.25	0.29	0.37
Roads	0.23	0.28	0.49
Construction	0.58	0.62	0.73
Residential	0.39	0.43	0.51
Forest	0.60	0.60	0.69
Highways	0.58	0.58	0.68
Outdoor sports	0.74	0.74	0.83
International Waters	0.59	0.63	0.68
City Roads	0.51	0.55	0.61
Village	0.41	0.50	0.59
Park news	0.29	0.37	0.46
Corridor news	0.80	0.84	0.90
Parking news	0.74	0.78	0.83
Finance	0.71	0.75	0.82
Politics	0.71	0.71	0.78
Local	0.67	0.69	0.79
Technology	0.44	0.52	0.57
Indoor sports	0.71	0.74	0.83

Table 6 AUC values for different news category

News category	Avg. AUC [NB]	Avg. AUC [SVM]	Avg. AUC [Proposed]
Agriculture	0.65	0.69	0.78
Aviation	0.37	0.46	0.63
Sports	0.60	0.66	0.79
Leisure	0.26	0.30	0.38
Roads	0.23	0.29	0.51
Construction	0.60	0.64	0.75
Residential	0.40	0.44	0.53
Forest	0.62	0.62	0.71
Highways	0.60	0.60	0.70
Outdoor sports	0.77	0.77	0.86
International Waters	0.61	0.65	0.70
City Roads	0.53	0.57	0.63
Village	0.42	0.52	0.61
Park news	0.30	0.38	0.47
Corridor news	0.83	0.87	0.93
Parking news	0.76	0.81	0.86
Finance	0.73	0.78	0.85
Politics	0.73	0.73	0.81
Local	0.69	0.71	0.82
Technology	0.45	0.54	0.59
Indoor sports	0.73	0.77	0.86

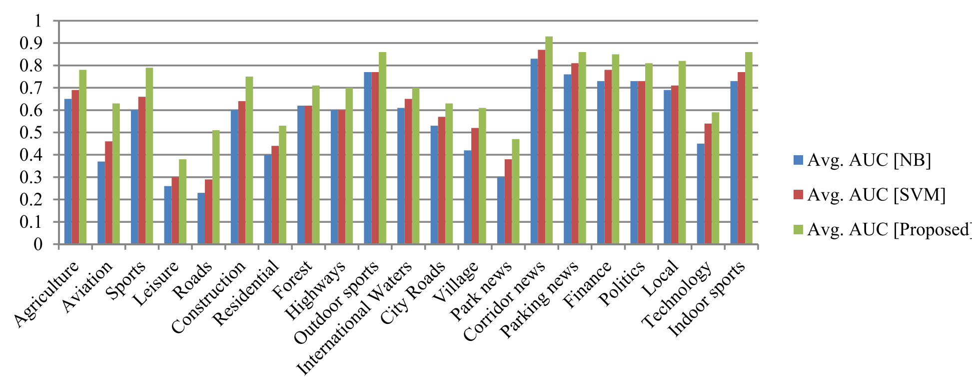

Fig. 7 Representation of AUC values for different news categories

future work, we aim to use more number of datasets, with more number of labels. We can also include use of special characters, numeric values and emoticons as well in the body of the news articles.

Acknowledgements The authors acknowledge the support given by the faculties and staff of the Department of Computer Science & Engineering, Biju Patnaik University of Technology, Rourkela

Author contributions All the authors have equally contributed to preparing the manuscript.

Funding Not applicable.

Availability of data and material Not applicable.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

Ahmad I, Yousaf M, Yousaf S, Ahmad MO (2020) Fake news detection using machine learning ensemble methods. Complexity. https://doi.org/10.1155/2020/8885861

Ali H et al (2021) All your fake detector are belong to us: evaluating adversarial robustness of fake-news detectors under black-box settings. IEEE Access. https://doi.org/10.1109/ACCESS.2021.3085875

Asghar MZ, Habib A, Habib A, Khan A, Ali R, Khattak A (2021) Exploring deep neural networks for rumor detection. J Ambient Intell Human Comput 12(4):4315–4333

Aslam N, Khan IU, Alotaibi FS, Aldaej LA, Aldubaikil AK (2021) Fake detect: a deep learning ensemble model for fake news detection. Complexity. https://doi.org/10.1155/2021/5557784

Chauhan P, Sharma N, Sikka G (2021) The emergence of social media data and sentiment analysis in election prediction. J Ambient Intell Human Comput 12(2):2601–2627

De Oliveira NR, Medeiros, DSV, Mattos DMF (2020) A sensitive stylistic approach to identify fake news on social networking. IEEE Signal Proc Lett 27:1250–1254. https://doi.org/10.1109/LSP.2020.3008087

Dong X, Victor U, Qian L (2020) Two-path deep semisupervised learning for timely fake news detection. IEEE Trans Comput Soc Syst 7(6):1386–1398

Dutta HS, Chakraborty T (2020) Blackmarket-driven collusion among retweeters-analysis, detection, and characterization. IEEE Trans Inf Forensics Secur 15:1935–1944. https://doi.org/10.1109/TIFS.2019.2953331

Elhadad MK, Li KF, Gebali F (2020) Detecting misleading information on COVID-19. IEEE Access 8:165201–165215. https://doi.org/10.1109/ACCESS.2020.3022867

Hakak S, Alazab M, Khan S, Gadekallu TR, Maddikunta PKR, Khan WZ (2021) An ensemble machine learning approach through effective feature extraction to classify fake news. Future Gener Comput Syst 117:47–58. https://doi.org/10.1016/j.future.2020.11.022

Han B, Han X, Zhang H, Li J, Cao X (2021) Fighting fake news: two stream network for deepfake detection via learnable SRM. IEEE Trans Biomet Behav Identify Sci 3(3):320–331

Henrique P, Faustini A, Covões TF (2020) Fake news detection in multiple platforms and languages. Exp Syst Appl. https://doi.org/10.1016/j.eswa.2020.113503

Jiang T, Li JP, Haq AU, Saboor A, Ali A (2021) A novel stacking approach for accurate detection of fake news. IEEE Access 9:22626–22639. https://doi.org/10.1109/ACCESS.2021.3056079

Kaliyar RK, Goswami A, Narang P (2021a) FakeBERT: fake news detection in social media with a BERT-based deep learning approach. Multimed Tools Appl 80(8):11765–11788

Kaliyar R K, Goswami A, Narang P (2021b) A hybrid model for effective fake news detection with a novel COVID-19 dataset. In: International Conference on agents and artificial intelligence (ICAART-2021b) 2: 1066–1072. https://doi.org/10.5220/0010316010661072

Khan W, Ghazanfar MA, Azam MA, Karami A, Alyoubi KH, Alfakeeh AS (2020) Stock market prediction using machine learning classifiers and social media, news. J Ambient Intell Human Comput. https://doi.org/10.1007/s12652-020-01839-w

Kumar S, Asthana R, Upadhyay S, Upreti N, Akbar M (2019) Fake news detection using deep learning models: a novel approach. Trans Emerging Tel Tech. https://doi.org/10.1002/ett.3767

Li Q, Hu Q, Lu Y, Yang Y, Cheng J (2019) Multi-level word features based on CNN for fake news detection in cultural communication. Pers Ubiquit Comput. https://doi.org/10.1007/s00779-019-01289-y

Li D, Guo H, Wang Z, Zheng Z (2021) Unsupervised fake news detection based on autoencoder. IEEE Acc 9:29356–29365. https://doi.org/10.1109/ACCESS.2021.3058809

Liu Y, Wu YB (2020) FNED: a deep network for fake news early detection on social media. ACM Trans Inf Syst. https://doi.org/10.1145/3386253

Mehta D, Dwivedi A, Patra A, Anand Kumar M (2021) A transformer-based architecture for fake news classification. Soc Net Anal Min. https://doi.org/10.1007/s13278-021-00738-y

Neves JC et al (2020) GANprintR: improved fakes and evaluation of the state of the art in face manipulation detection. IEEE J Sel Top Signal Proc 14(5):1038–1048

Ni S, Li J, Kao HY (2021) MVAN: multi-view attention networks for fake news detection on social media. IEEE Acc 9:106907–106917. https://doi.org/10.1109/ACCESS.2021.3100245

Raj C, Meel P (2021) ConvNet frameworks for multi-modal fake news detection. Appl Intell 51(11):8132–8148. https://doi.org/10.1007/s10489-021-02345-y

Sahoo SR, Gupta BB (2021) Multiple features based approach for automatic fake news detection on social networks using deep learning. Appl Soft Comput 100:106983. https://doi.org/10.1016/j.asoc.2020.106983

Sansonetti G, Gasparetti F, D'Aniello G, Micarelli A (2020) Unreliable users detection in social media: deep learning techniques for automatic detection. IEEE Acc 8:213154–213167. https://doi.org/10.1109/ACCESS.2020.3040604

Shim JS, Lee Y, Ahn H (2021) A link2vec-based fake news detection model using web search results. Exp Syst Appl. https://doi.org/10.1016/j.eswa.2021.115491

Shrivastava G, Kumar P, Ojha RP, Srivastava PK, Mohan S, Srivastava G (2020) Defensive modeling of fake news through online social networks. IEEE Trans Comp Social Syst 7(5):1159–1167

Subbanarasimha RP, Srinivasa S, Mandyam S (2020) Invisible stories that drive online social cognition. IEEE Trans Comput Soc Syst 7(5):1264–1277. https://doi.org/10.1109/TCSS.2020.3009474

Thota A, Tilak P, Ahluwalia S, Lohia N (2018) Fake news detection: a deep learning approach. SMU Data Sci Rev 1(3):10. https://scholar.smu.edu/datasciencereview/vol1/iss3/10

Umer M, Imtiaz Z, Ullah S, Mehmood A, Choi GS, On BW (2020) Fake news stance detection using deep learning architecture (CNN-LSTM). IEEE Access 8:156695–156706. https://doi.org/10.1109/ACCESS.2020.3019735

Varshney D, Vishwakarma DK (2021) Hoax news-inspector: a real-time prediction of fake news using content resemblance over web search results for authenticating the credibility of news articles. J Ambient Intell Human Comput 12(9):8961–8974. https://doi.org/10.1007/s12652-020-02698-1

Verma PK, Agrawal P, Amorim I, Prodan R (2021) WELFake: word embedding over linguistic features for fake news detection. IEEE Trans Comput Soc Syst 8(4):881–893. https://doi.org/10.1109/TCSS.2021.3068519

Xu K, Wang F, Wang H, Yang B (2020) Detecting fake news over online social media via domain reputations and content understanding. Tsinghua Sci Tech 25(1):20–27. https://doi.org/10.26599/TST.2018.9010139