Multimedia Tools and Applications (2024) 83:28579–28594
https://doi.org/10.1007/s11042-023-16588-z

FakeIDCA: Fake news detection with incremental deep learning based concept drift adaption

Shubhangi Suryawanshi[1,2] · **Anurag Goswami**[1] · **Pramod Patil**[2]

Received: 6 December 2022 / Revised: 2 August 2023 / Accepted: 21 August 2023 /
Published online: 6 September 2023

Abstract

Social media facilitates rapid information sharing, improving exposure, connections, and content promotion. However, it also poses the challenge of fake news, which can mislead and harm individuals physically, and mentally, and incite violence. Fake news is often known as incorrect or misleading information. Prior research used Machine Learning (ML) and Deep Learning (DL) techniques for fake news detection. These studies predominantly relied on static offline models, overlooking the dynamic and evolving nature of news patterns, assuming their stability over time. Our paper proposes an incremental ensemble neural network for fake news detection that continuously learns from fake news streams, adapting to changes. It employs performance-based pruning to eliminate underperforming classifiers, improving overall performance. Additionally, the model detects concept drift in real-time and triggers adaptation strategies to maintain accuracy and robustness. The models undergo evaluation in two scenarios, utilizing consistent news patterns for training and testing, demonstrating consistent performance among all ML and incremental models. In the second scenario, the study analyzes the impact of news patterns over time, including concept drift due to significant events like the United States election. The analysis reveals that offline-trained methods are susceptible to performance degradation. However, the proposed model exhibits consistent performance with an accuracy of 97.90% and 99.76% on two fake news datasets, despite changes in the news pattern over time. The findings demonstrate how the evolution of the news pattern impacts the effectiveness of fake news detection models. The proposed model used for the experimentation indicates consistent performance even in the presence of drift.

Keywords Fake news · Deep learning · Incremental learning · Concept drift

✉ Shubhangi Suryawanshi
ss5683@bennett.edu.in

Anurag Goswami
anurag.goswami@bennett.edu.in

Pramod Patil
pdpatiljune@gmail.com

[1] Bennett University, Greater Noida, India

[2] Dr. D. Y. Patil Institute of Technology, Pune, India

1 Introduction

The modern era of technological advancements has witnessed a rapid increase in the number of users of digital media. Information is accessible with a single click via handheld devices (e.g., smartphones). However, this ease of sharing news has led to rapid growth in the spread of false information, which has been used to influence the decision of users socially, politically, and economically. Fake news [1, 2], disguised as genuine information, can deceive readers, causing misinformation and manipulation of public opinion. Its impact extends beyond spreading falsehoods, potentially harming individuals' well-being and inciting violence [1–3]. During critical events like elections [1], the influence of fake news can significantly sway voters' decisions and affect democratic processes. Due to the popularity of social media and the ability, it offers users to create accounts anonymously, the number of malicious users has increased [31]. This has, among other issues, also resulted in the dissemination of biased or incorrect information relating to health, politics, and business. Consequently, user perceptions of these subjects have been wrongly influenced. Such information also has an impact on real-world events such as elections [1], health-related problems [1, 2], business [1], and conspiracy theories about the emerging coronavirus (COVID-19) pandemic [1, 4, 5].

The extensive use of popular social networking sites such as Facebook, Twitter, Instagram, and Whatsapp, along with traditional media hugely increases disinformation and its spread without verification for correctness [26, 29, 31, 32]. The COVID-19 infodemic (i.e., false or misleading information in media either digital or physical environment) quickly spread across the globe during the pandemic [WHO] [5, 30]. It misled users' opinions and created confusion among users, tremendously increasing risk-taking behaviour, and resulting in an adverse impact on health. With the expansion of social media due to growing digitization, false information spreads rapidly. As such, there is a need for fake news detection. While information correctness or accuracy can be assessed by a fact-checker, the volume of information generated every day is massive and manual checking of this information can be a tedious job. A significant number of existing research have failed to address the time-sensitive aspect of news, specifically how news evolves over time or the recurrence of certain news patterns. Additionally, these studies have predominantly used offline learning techniques [1, 3, 4, 13–17, 27, 28].

Several researchers have used machine learning and deep learning techniques to automatically detect fake news. Previous studies [6, 9–11, 13, 15, 17, 18] considered the data to be static. They presumed that the data never evolve, and hence, employed offline learning. Machine learning and Deep learning approaches are the leading approaches for fake news detection; however, they do not take into account the evolving nature of the features or the target label and employ offline learning to detect and analyze fake news. As fake news contains a continuous stream of data (i.e., news) [19], it may evolve over a period of time [16, 23]. External events (e.g., US elections [29], COVID-19 [30]) cause news topics to shift substantially over time. As such, the relationship between content and target labels changes as well. Concept drift occurs when the distribution of input data or the relationship between input data and target label changes over time [24, 25]; this impacts the performance of a previously trained model. Offline learning does not permit incremental model training [8], which overlooks the possibility that news patterns may evolve over time; hence, the model produces inflated results [25, 26]

Addressing this pressing issue has become imperative in the interconnected world, where social media plays a central role. However, the dynamic and continuously evolving nature of fake news data poses a substantial challenge to existing static models. Traditional machine learning and deep learning techniques [6, 9–11, 13, 15, 17, 18], which are trained on fixed datasets, struggle to adapt to changes in data patterns over time. As fake news creators continually evolve their tactics and narratives, static models may become less accurate, compromising their effectiveness in real-time detection. The research is motivated by the alarming rise of evolving fake news in the digital age, facilitated by social media and online platforms, leading to widespread dissemination in mere minutes.

To effectively combat fake news, there is a critical need for innovative and adaptive approaches that can continuously learn and adapt to the ever-changing fake news landscape. The incremental learning method can be used to solve the problem of evolving fake news, as it adapts to new data patterns quickly [7, 37].

To address the aforementioned problem, The study is proposed with the primary goal to develop an incremental ensemble neural network model that addresses the challenges posed by concept drift in fake news detection. By continuously learning and updating its knowledge to detect emerging fake news patterns, the proposed model aims to maintain high accuracy and reliability in detecting fake news over time. The objective is to build a robust and effective fake news detection system that remains accurate and reliable over time, even as the nature of fake news data evolves.

The contribution of this paper is as follows:

– Proposes an incremental ensemble neural network an adaptive model for fake news detection that continuously learns and adapts to changes in the fake news data stream. Unlike traditional static models, this model dynamically updates its knowledge based on new data, improving its ability to detect fake news.
– Introduces a performance-based pruning technique that selectively removes underperforming classifiers from the ensemble. This enhances the overall performance by eliminating less useful member.
– Incorporates drift detection mechanisms to identify when concept drift occurs and triggers adaptation strategies to update the ensemble accordingly.
– Investigates and analyzes the impact of concept drift on the performance of the fake news detection model. Provides insights into how concept drift affects the accuracy, stability, and generalization capabilities of the model, contributing to the understanding of the challenges posed by evolving data patterns.
– The evaluation and comparison of the proposed model with existing machine learning and incremental learning techniques using a fake news datasets.

The paper is organized in the following manner. Section 2 describes works related to concept drift detection methods and fake news detection models. Section 3 describes the methodology with detailed illustrations. Section 4 presents details of the experiment and a comparison of the existing methods and proposed model on the fake news dataset. Finally, Section 5 concludes this work and discusses the possible future scope of this work.

2 Related work

In the past studies [6, 10, 11, 13, 27, 28], techniques to identify fake news have primarily focused on content. Content-based methods have been used to build fake news detection models using a variety of fake news data, including article content, new source, and headline. Context-based methods analyse the user's social behaviour [33–35] such as posts, comments, and responses, as well as network relationship elements, such as followers-followee. For the categorization of fake news, features based on content are collected from the visual or textual contents of the news [8]. Several approaches, including Term Frequency and Inverse Document Frequency (TF-IDF) [20], Bag-of-words [21], Word2Vec [22], and Global Vectors for Word Representation (GloVe) [21] were employed to extract features in fake news detection. To detect fake news, support vector machines [10], logistic regression [10–13], random forest [11–13], naive Bayes [11, 12], decision tree (DT), stochastic gradient descent (SGD) [6, 10–13], long short-term memory (LSTM) [6], and gated recurrent unit (GRU) [15] were utilized.

A study [11, 12] used a variety of machine learning techniques, including Term Frequency and Inverse Document Frequency for feature extraction, wherein the linear regression-based model achieved 89% accuracy and the linear support vector machine classifier achieved 92% accuracy. Several studies have employed the deep neural network model as well [6, 15, 27, 28]. A study [14] presented a deep neural network that detected fake news with 48.80% accuracy. One study [15] used a hybrid model to classify fake news and achieved an accuracy of 89.20%.

The study [16] employed the convolution neural network (CNN) and bi-directional LSTM. The output of both models fed to the multilayer perception for the final result. The study used an ensemble of CNN and Bi-LSTM networks. The multi-layer perceptron took the output from the entire ensemble network to produce the final result. This approach achieved 44.87% accuracy. The results of this study were compared with various deep learning approaches, including LSTM, hybrid CNN, and CNN models. In addition, many ensemble approaches for fake news detection have been presented, including random forest, which consists of an ensemble of decision trees [17]. Ensemble approaches are useful in minimizing generalization errors and enhancing overall performance. An ensemble approach with varied textual features [18] and other machine learning algorithms [10–13] were employed in an ensemble for fake news detection.

Another study [24, 25] investigated the effect of concept drift on the classification of fake news. The results reveal that the performance of the model trained offline diminishes with time as news pattern changes. The findings of the study show that the offline-trained model can be overly optimistic and that it is preferable to use incremental learning to adapt to changes over time [25].

The study [38] introduces an innovative method for multimedia news summarization in internet search results. The proposed approach leverages the hierarchical latent Dirichlet allocation (hLDA) model to uncover the hierarchical topic structure within news documents relevant to a given query. Additionally, it employs a time-bias maximum spanning tree (MST) algorithm to interconnect these documents and generate a concise and cohesive summary representing the overarching topic.

The study [39] presents an understanding-oriented approach to multimedia news summarization for internet search results, tackling the challenge of handling massive multimedia data. By uncovering underlying topics and threading news events within each topic, the method generates concise overviews for users. Additionally, the paper introduces VizByWiki, a system that enhances news articles with relevant data visualizations from Wikimedia Commons, which achieved a successful augmentation rate. The system can also rank visualizations by their usefulness, providing a valuable resource for readers and researchers.

As mentioned earlier, several researchers have applied a variety of machine learning techniques without considering the dynamic nature of the features or target label, which has resulted in the methods being outperformed in some datasets [23–25] and performing poorly on others [23, 24]. The offline model fails to consider the temporal patterns of news (changes in news patterns over time) and gives inflated results, as it is not trained gradually. Some research was undertaken to examine the impact of time on fake news [23–25]. The impact of time on the classifier employed for fake news identification was examined in [23–25]. The results reveal that the classifier's performance deteriorates over time. This problem can be solved through online learning or incremental learning [7, 37]. The study of concept drift on fake review detection was presented in a research article [24]. The study [24] employed the drift detection method (DDM), and adaptive windowing (ADWIN) approaches. ADWIN approach uses a sliding window of data to detect concept drifts. The two dynamically changing sub-windows, which stand for both the recent and historical data, are saved. The oldest data is removed from the adjustable window when the mean values of two sub-windows are noticeably greater than a particular threshold, and a drift is detected. The experiment was carried out by employing a drift detection approach and a classification method based on content to analyze the correlation between concept drift and model performance. The findings revealed that concept drift had a negative impact on the classification model [24, 25].

The research community is interested in fake news detection because it has become a significant topic in the real world. Although machine learning and deep learning models have received a lot of attention, as mentioned earlier, they don't take into account the changing nature of news and how concept drift can impair the learning models' accuracy over time. As such, a learning model's performance may be hindered if the model trained on one dataset is unable to identify new news patterns that emerge over time [23–25].

The various techniques used for fake news detection, namely content-based and context-based methods [33–35], as well as the utilization of different machine learning and deep learning algorithms [6, 10–13, 15, 27, 28]. Content-based approaches focus on features extracted from the textual or visual content of news articles using methods like TF-IDF [20], Bag-of-words [21], Word2Vec [22], and GloVe [21]. On the other hand, context-based methods analyze users' social behavior, including posts, comments, and responses, to detect fake news. The studies have achieved varying levels of accuracy using different techniques. For instance, some studies achieved high accuracy using linear regression-based models or linear support vector machine classifiers. Deep neural network models and hybrid approaches have also demonstrated promising results. Ensemble methods, such as using CNN [16, 27] and Bi-LSTM [6] networks in combination with multi-layer perceptron, have been employed to enhance performance and minimize generalization errors.

Furthermore, as observed in several studies, the ensemble approach [16–18, 24] offers promising results when compared to single models, benefiting from collective intelligence and minimising generalisation errors. A major problem identified in these studies is the failure to recognise and adapt to the dynamic nature of fake news data and the impact of concept drift over time. As news trends evolved over time, the offline-trained models' performance deteriorated, resulting in inflated results and decreased accuracy. This problem emphasises the importance of an incremental learning strategy that can adapt to changing data patterns while maintaining performance over time.

The literature review emphasises the challenges and limitations of current fake news detection methods. It emphasises the significance of concept drift and the need for an incremental learning approach. The review of the literature gives useful insights into the issues faced by existing fake news detection algorithms and emphasises the need for dynamic adaptation in addressing concept drift.

The challenges and limitations of existing methods for detecting fake news motivated us to explore more accurate and effective solutions in the booming field of evolving fake news detection. The proposed model addresses the limitations observed in existing studies by employing an incremental ensemble approach to handle concept drift and adapt to changes in fake news data over time. The proposed model ensures real-time adaptability that effectively maintains accuracy despite changing data patterns by employing incremental learning. Its ability to adapt in real-time to changing data patterns increases its effectiveness in handling and detecting evolving fake news patterns. The proposed model in the present study offers a promising solution by addressing these challenges and providing a more robust and efficient solution for combating the spread of fake news in real-world scenarios.

3 Experiment design

This section details the research goal, the dataset used for the experimentation, and the methodology followed. The proposed model (described in Section 3.4) is evaluated using fake news datasets and analyzes the impact of concept drift on the fake news detection model.

3.1 Research goal

To investigate and evaluate the effect of the evolving nature of fake news on the performance of an incremental deep learning based fake news detection models using the concept drift technique.

3.2 Dataset

The details of the datasets utilized for the experiments are as follows:

– The fake news dataset [https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset] used for the experimentation contains a total of 40525 instances with title, text, subject, and date as features and the target label contains the value fake or not fake.
– The "Getting Real about Fake News" dataset was used for the experimentation. Originating from the user-generated content of the BS Detector Chrome Extension, the dataset contains 36,300 instances. Each instance within this collection consists of a title, the article's content, timestamp, and a target label.

3.3 Evaluation measures

The evaluation metrics for the experiments are described as follows:

– **Accuracy** - is calculated by dividing the total number of correct predictions by the total number of predictions.

$$\text{Accuracy} = \text{Correct\_Prediction}/\ \text{Total\_Prediction}$$

– **F1-score** - The harmonic mean of precision and recall is called the F1 score.
– **Precision** - is calculated as the ratio of positive data samples that were correctly classified to all samples that were classified as positive.
– **Recall** - is determined as the proportion of positive samples that were correctly identified as positive to all positive samples.

3.4 Methodology

Figure 1 depicts the workflow for the proposed model. The processes include generating stream data (1), pre-processing (2), feature extraction using TF-IDF (3), detecting drift (4),

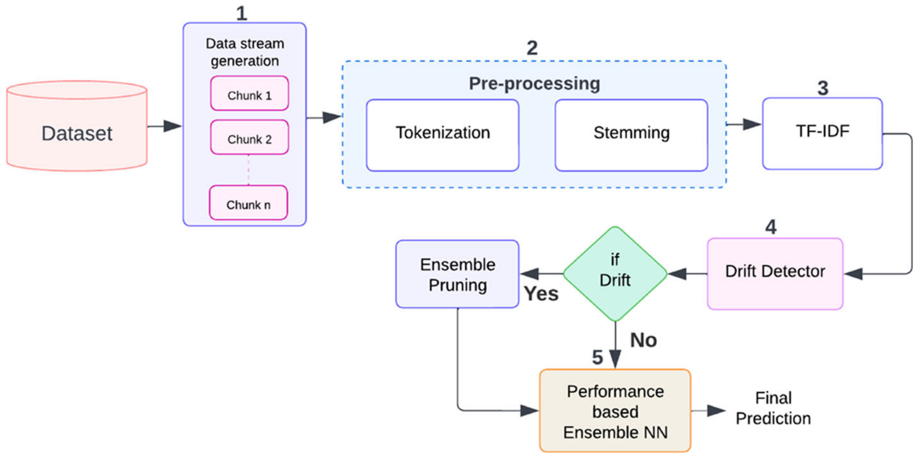

Fig. 1 Fake News Detection with Incremental Deep Learning Based Concept Drift Adaption

and using a performance-based incremental ensemble classifier (5). The model has been assessed using fake news datasets, which is split into two parts: D1 and D2, for analysis. State-of-the-art machine learning techniques and an incremental model were used to evaluate the proposed model by considering its accuracy, precision, recall, and F1 score.

The steps involved in the methodology are described in detail below:

Step 1: Stream Data Generation - Two datasets are utilized for an experiment, both partitioned into D1 and D2. Each dataset is divided into portions, corresponding to different periods. The first dataset includes 18,720 instances from 2015 to 2016 in D1 and 21,805 instances from 2017 to 2018 in D2. Each instance in this dataset is characterized by features such as title, text, subject, and date, and also includes a target label to indicate if the news is real or fake. The second dataset is divided into instances from 2013 to 2015 in D1 and from 2016 to 2018 in D2, with a total of 36,300 instances. These instances include data attributes like title, content, and date, along with the target label to indicate if the news is real or fake. Additionally, the dataset is segmented into equal-sized chunks, forming a data stream for further processing.

Step 2: Data Pre-Processing - The data is cleaned by eliminating punctuation marks and special symbols that might not contribute to the meaningful interpretation of the text. This is followed by tokenization, a process where the text is broken down into individual words or 'tokens'. Stemming comes next, where each word is reduced to its root form, ensuring consistent representation and preventing redundancy. To ensure uniformity and eliminate potential duplication based on case differences, all the words are converted to lowercase. In the subsequent stage, stop words - common words that often do not carry significant meaning - are discarded. Lastly, N-grams are employed, an approach that considers the occurrence of 'n' consecutive words, to capture contextual information and enhance the depth of text analysis.

Step 3: Term Frequency and Inverse Document Frequency (TF-IDF) - In this step, TF-IDF along with N-gram is used for feature extraction to identify important words based on their occurrence in the text. N-gram helps preserve the word order, providing context for the extracted features.

Term Frequency - This step focuses on calculating the frequency of each word in the text. It counts how often each word appears in a given text. This frequency is then divided by the total number of words present in that text. This is represented as follows:

$$\text{Term Frequency} = C_{word}/Total_{word}$$

The C_{word} represents the count of words in the text and $Total_{word}$ represents a total number of words in the text.

Inverse Document Frequency(IDF) - While term frequency measures the occurrence of a word in a text, IDF gauges how important that word is in the entire corpus. IDF is calculated using the formula:

Inverse Document Frequency = log(N/Number of texts containing the word)

Here, 'N' represents the total number of texts in the dataset.

The weight of each word in a given text is calculated. This is shown as follows:

$$Weight_{word,text} = TermFrequency_{word,text} * log(N/text_{word}) \tag{1}$$

Term $Frequency_{word,text}$ shows the count of words in a given text. $Weight_{word,text}$ reflects the weight of the word in the given text (shows importance or weight of the word). Lastly, N is the number of titles in the fake news dataset.

Step 4: Drift Detector - The Page-Hinkley (PH) test is used for detecting drifts (change points) in a sequence of observations [36].

- Set the initial monitoring variables cumulative sum of differences D_c and the minimum of these sums D_m to zero.
- Calculate the mean μ of the observed values up to the current point in time. This mean μ will be updated for each new data point, the difference between the data point and the mean is calculated as shown in (2).

$$D_c(t) = D_c(t-1) + (x(t) - \mu(t)) \tag{2}$$

  where x(t) is the new observation at time t, μ(t) is the mean of the observations up to time t, and D_c(t-1) is the previous cumulative sum.
- After updating D_c, update D_m as shown in (3).

$$D_m(t) = min(D_m(t-1), D_c(t)) \tag{3}$$

- Compute the Page-Hinkley statistic as shown in (4).

$$PH(t) = D_c(t) - D_m(t) \tag{4}$$

- If PH(t) exceeds a predefined threshold λ, it indicates a drift or a change point has occurred. The λ value is taken into account as the threshold, which is determined by considering the allowable false alarm rate. As the threshold value increases, the false alarm rate reduces while the likelihood of delaying the changes increases.

The Page-Hinkley signals when a drift has occurred. If a drift is detected, the classifier's performance is validated, and ensemble pruning is carried out based on its real-time error rate.

Step 5: Performance-based Ensemble Classifier - The steps involved in the Performance-based Ensemble Classifier are as follows:

- The model includes an ensemble of four distinct neural networks, represented as E=N1, N2, N3, N4. Each of these neural networks acts as a separate model within the ensemble.

- The true label of each instance is denoted by Test[Y], while the label predicted by the model is represented by TestPredict.
- The real-time error RT_{error} is calculated as the difference between the true label and the predicted label. If RT_{error} equals zero, it means the model has correctly classified the instance.
- When the classification is correct (RT_{error}==0), the weight 'w' of the classifier is set to 1. However, when the drift is detected, the real-time error rate of each neural network model is calculated.

$$RT_{error} = Test[Y] - TestPredict[i] \tag{5}$$

- The weight for each classifier within the ensemble is calculated based on its real-time error rate as shown in (6).

$$w_i = \sum_{i=1}^{n} 1/RT_{error} \tag{6}$$

- The final prediction (Pred) of the ensemble is determined by summing the products of the weights and the predictions from each classifier within the ensemble as shown in (7).

$$Pred = \sum_{i=1}^{n} w_i * TestPredict[i] \tag{7}$$

- The ensemble is designed to continually improve itself by calculating the real-time error rate of each of its members when the drift is detected. Any underperforming member, detected by its higher error rate, is removed from the ensemble.
- New member (neural network model) trained on the most recent data instances and added to the ensemble. This ensures that the ensemble stays updated and adaptive to any evolving data patterns.

4 Result and discussion

The experiment was performed using two fake news datasets. The analysis of the proposed model and the impact of concept drift on model performance was the prime purpose of dataset division (into D1 and D2). A prequential approach was used for this experiment. A prequential approach initially tests the data before using it for training. Python, Sklearn, and River online machine learning library were utilized for the implementation. Machine learning techniques such as Naive Bayes, Decision Trees, SVM [24], Multi-layer Perceptron [19], and Hoeffding Tree [19], OzaBagging [19, 24] and Passive Aggressive Classifier [37] and Adaptive Random Forest incremental learning models were used for comparison.
The experiment setup was as follows:

The proposed model consists of four different neural network models in an ensemble. In the basic setup of the network, there were two dense layers with 128 and 64 neurons each, and the output layer of a neural network is comprised of one neuron with sigmoid activation. The binary cross-entropy loss function was used, the learning rate was set to 0.01 and the Adam optimizer was employed. The model was trained in a single pass.

The drift detection was performed using the Page-Hinkley statistical test. The Page-Hinkley statistical test's parameters λ and δ were set to 50 and 0.005, respectively. The weighting factor δ is used to balance the observed value and the mean is called the forgetting factor and λ is the threshold of the change detection. The permissible false alarm rate is taken into account while determining the threshold value. While raising the threshold value lowers false alarms, it also raises the possibility of missing or postponing changes. Since change detection is a critical aspect of this study, the threshold value was set at 50 to achieve a balance between the likelihood of false alarms and the risk of missing or delaying changes.

The experiments were performed to assess the proposed model's ability to adapt to changing news patterns. Through these experiments, a comprehensive comparison was made between the proposed model and various other models, to analyze their respective abilities to handle dynamic data distributions and their overall effectiveness in detecting fake news in real-world scenarios. The evaluation process consisted of two scenarios and utilized two datasets. To understand the effects of changing news patterns, the dataset was divided based on significant events like the US election. News from before the election was assigned to D1, while news from after the election was included in D2. A similar approach was taken with the second dataset, where news was distributed between D1 (from 2013 to 2015) and D2 (from 2016 to 2018). This setup helped to investigate the models' adaptability to evolving news content over time.

4.1 Scenario1

In the first scenario, D1 dataset from both datasets was employed for incremental training and testing. Here, D1 was divided into 80% for training and 20% for testing. As shown in Table 1, In scenario 1, a comprehensive evaluation of various models was conducted using the Fake and Real news datasets, along with the 'Getting Real about Fake News' dataset. The results demonstrated that all the models, including machine learning models, incremental learning models, and the proposed model, achieved consistent accuracy when the same dataset D1 was utilized for both training and testing. This consistency in accuracy indicates that the models were able to generalize well to the data and maintain their performance levels when tested on the same part of the dataset used for training.

4.2 Scenario 2

In scenario 2, the D1 dataset was utilized for training while the D2 dataset was employed for testing. Conducted an extensive evaluation of various machine learning and incremental learning models, including Adaptive Random Forest, Oza Bagging, Passive Aggressive Classifier, and Hoeffding Tree, initially using Fake and Real news datasets. Our findings illustrated that these incremental learning models achieved superior accuracy (95.91%, 89.18%, 74.70% and 93.58% respectively) compared to traditional static models like Naive Bayes, Decision Tree, SVM, and Multi-layer Perceptron. Notably, our proposed model excelled, obtaining an accuracy of 97.90%, even when different datasets were used for training and testing, indicating its ability to adeptly handle the evolving nature of fake news.

Further experimentation was conducted on the 'Getting Real about Fake News' dataset to validate the efficiency of our proposed model, and to compare it with existing methods. Here, it was observed that static models such as Naive Bayes, Decision Tree, SVM, and Multi-layer Perceptron experienced a decline in performance when trained and tested on different datasets, underscoring their inadequacy in dealing with the changing landscape of fake news. On the contrary, incremental models like Adaptive Random Forest, Passive-Aggressive Classifier, Oza Bagging Classifier, and our proposed model maintained high accuracy rates (91.71%, 99.23%, 90.23% and 99.76% respectively), even as fake news patterns evolved over time. as shown in Table 1.

Table 1 Comparison of existing methods with the proposed model

Dataset		Model	Accuracy	F1 Score	Precision	Recall
Fake and real news dataset						
D1	D1 for Training and Testing	Naıve Bayes	97.34	93.47	92.0	94.25
		Decision Tree	97.64	93.50	93.36	93.64
		SVM [24]	97.54	95.14	95.23	94.02
		Multi-layer Perceptron [19, 37]	96.32	94.11	95.63	94.18
		Hoeffding Tree [19]	98.99	**97.92**	96.56	95.33
		Adaptive Random Forest	90.0	90.12	90.14	89.92
		Passive-Aggressive Classifier [37]	97.44	97.24	**97.38**	**97.10**
		Ozabagging Classifier [24]	94.32	89.17	90.62	91.62
		Proposed Model	**97.92**	96.37	95.68	96.33
D1-D2	D1 for Training and D2 for Testing	Naıve Bayes	45.55	41.75	61.16	55.09
		Decision Tree	36.98	28.61	68.05	51.06
		SVM [24]	48.12	44.75	59.89	52.70
		Multi-layer Perceptron [19, 37]	42.54	44.53	51.23	48.12
		Hoeffding Tree [19]	93.58	91.68	92.44	91.92
		Adaptive Random Forest	95.91	90.92	90.94	90.92
		Passive-Aggressive Classifier [37]	74.70	54.56	54.26	50.18
		Ozabagging Classifier [24]	89.18	92.39	86.0	91.81
		Proposed Model	**97.90**	**93.37**	**94.68**	**92.33**
Getting Real about Fake news dataset						
D1	D1 for Training and Testing	Naıve Bayes	98.78	98.66	98.33	99.03
		Decision Tree	98.83	**98.90**	**99.87**	**99.80**
		SVM [24]	97.12	95.26	93.13	97.83
		Multi-layer Perceptron [19, 37]	81.8	75	76.9	73.1
		Hoeffding Tree [19]	62.5	60	62.3	58.2
		Adaptive Random Forest	95.78	92.85	96.37	94.58
		Passive-Aggressive Classifier [37]	99.1	98.87	98.82	97.32
		Ozabagging Classifier [24]	91.31	94.99	91.61	98.61
		Proposed Model	**99.22**	95.59	95.89	95.91
D1-D2	D1 for Training and D2 for Testing	Naıve Bayes	66.53	39.95	50.00	33.26
		Decision Tree	34.33	25.55	17.16	50.00
		SVM [24]	74.88	42.82	50.00	37.44
		Multi-layer Perceptron [19, 37]	56.1	47	45.6	48.2
		Hoeffding Tree [19]	46.1	43	41.9	44.8
		Adaptive Random Forest	91.71	95.1	93.95	96.27
		Passive-Aggressive Classifier [37]	99.23	49.80	50.00	49.61
		Ozabagging Classifier [24]	90.23	92.43	90.32	96.18
		Proposed Model	**99.76**	**98.45**	**95.78**	**94.90**

The bold entries correspond to the best results observed in the study

As demonstrated in Table 1, the performance of Naive Bayes, Decision Trees, Multi-layer Perceptron, and SVM declines significantly when D1 is used for training and D2 for testing on the Fake and Real news dataset with an accuracy of 45.55%, 36.98%, 42.54% and 48.12%, respectively,

The results demonstrated that the incremental models, including Adaptive Random Forest, Oza Bagging, Passive Aggressive Classifier and the Proposed Model, exhibited consistent performance even as the news patterns evolved. Notably, the Proposed Model outperformed other state-of-the-art machine learning and incremental learning models. These findings highlight the effectiveness of the proposed model in handling concept drift and adapting to changing data distributions. This analysis shows the effectiveness of our proposed model in addressing the challenges posed by evolving fake news data.

The proposed model's confusion matrix demonstrates that the model accurately anticipates whether the data is truly positive or truly negative as compared to conventional machine learning approaches and incremental learning methods. The confusion matrix for the proposed model is shown in Figs. 2 and 3, which exhibits better true positive and true negative values than existing techniques on both datasets using different datasets for training and testing. It demonstrates that the model retains its accuracy across different datasets used for training and testing, indicating its ability to effectively combat the constantly evolving face of fake news.

The results indicate that the model performs well if it is trained and tested on the same dataset. As the same dataset was used for both training and testing in all previous experiments [6, 8, 9, 13, 15, 16, 18], the model's performance was maintained. In contrast, fake news detection with incremental deep learning based concept drift adaption works well and maintains the model's performance even if drift occurs as the ensemble model learns the data incrementally, adding the benefits of ensemble learning as well, which combines the predictions from various neural networks to reduce the generalization and prediction errors.

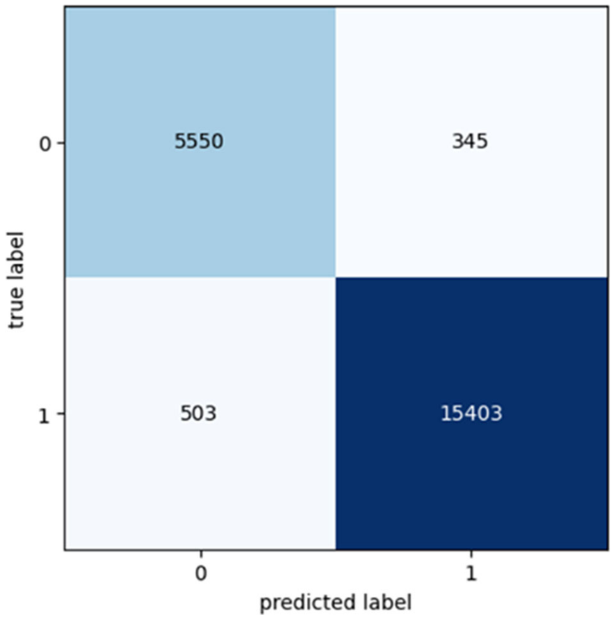

Fig. 2 Confusion Matrix for Proposed Model for Fake and Real news dataset

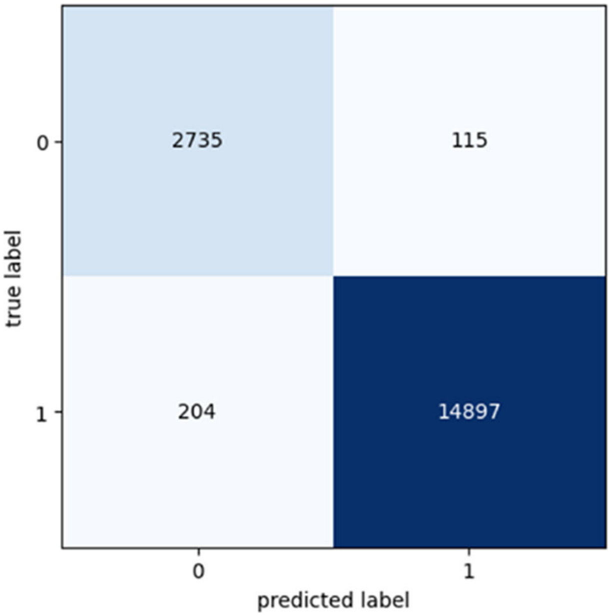

Fig. 3 Confusion Matrix for Proposed Model for Getting Real about Fake news dataset

However, the results of scenario 2 revealed that the traditional ML models become obsolete over time and their performance degrades. The results of the experiment show that concept drift and fake news detection have a negative correlation, which reduces the effectiveness of the fake news detection model.

5 Conclusion

In this study, we proposed a novel approach for detecting fake news employing incremental deep learning to manage concept drift adaptation, thereby enhancing the fake news detection process. Two fake news datasets were used in our experiment and divided into D1 and D2 for the analysis. We compared our proposed model's performance with several state-of-the-art machine learning and incremental learning methods, including Naive Bayes, Decision Trees, Multi-layer Perceptron, SVM, Adaptive Random Forest, Oza Bagging, Passive Aggressive Classifier, and Hoeffding Tree, and evaluated them based on accuracy, F1-score, precision, and recall. For the first scenario, the same dataset (D1) was used for both training and testing. However, in the second scenario, we used D1 for training and D2 for testing.

The results showed that offline models' performance deteriorated over time as news patterns changed, highlighting the limitation of assuming static characteristics for fake news data. In contrast, our proposed model, which employed an incremental ensemble neural network, continuously learned from the stream of fake news and adapted to changes over time. This model featured performance-based pruning to discard underperforming classifiers, thereby enhancing overall performance. Notably, the model was capable of identifying concept drift in real-time and activated adaptation strategies accordingly to preserve its accuracy and robustness. The proposed model exhibits consistent performance with an accuracy of 97.90% and 99.76% on two fake news datasets, despite changes in the news pattern over time. This illustrates the model's ability to effectively deal with the dynamic nature of fake news.

As our experiment indicated, the performance of traditional fake news detection techniques reduces as the concept drift increases due to the evolution of fake news. However, our proposed model using incremental deep learning for concept drift adaptation effectively maintains its performance, demonstrating its ability to adapt to the evolving nature of fake news, providing a robust and effective solution for fake news detection in a rapidly changing digital environment.

Our research was confined to analyzing textual data classification algorithms solely for English-language fake news detection. To overcome these limitations, future research could investigate the effectiveness of the considered classifiers for handling multi-lingual context-based fake news detection and explore the applicability of class imbalance and feature extraction techniques. In the future, self-evolving neural networks can be applied for fake news detection with concept drift adaption.

Data Availability The dataset that we used, is available online and the link is provided in the article.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Stephan Lewandowsky, Ullrich K. H. Ecker, Colleen M. Seifert, Norbert Schwarz, and John Cook (2012) Misinformation and its correction: Continued influence and successful debiasing. Psychol Sci Public Interest 13, 3 (2012)
2. Ying Lin, Joe Hoover, Gwenyth Portillo-Wightman, Christina Park, Morteza Dehghani, and Heng Ji. (2018) Acquiring background knowledge to improve moral value prediction. In Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'18). IEEE, 552–559
3. Ramy Baly, Georgi Karadzhov, Dimitar Alexandrov, James Glass, and PreslavNakov (2018) Predicting factuality of reporting and bias of news media sources. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'18)
4. Ramy Baly, Georgi Karadzhov, Abdelrhman Saleh, James Glass, and PreslavNakov (2019) Multi-task ordinal regression for jointly predicting the trustworthiness and the leading political ideology of news media. arXiv preprint arXiv:1904.00542 (2019)
5. Galhardi CP, Freire NP, Minayo MC. de S., Fagundes MCM (2020) Fato ou Fake? Uma análise da desinformaçãofrente à pandemia da Covid-19 no Brasil.-4210 https://doi.org/10.1590/1413-812320202510.2.28922020
6. Alves JL, Weitzel L, Quaresma P, Cardoso CE, Cunha L (2019) Brazilian presidential elections in the era of misinformation: A machine learning approach to analyse fake news. In Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, I. Nyström, Y. Hernández Heredia, and V. Milián Núñez (Eds.). Springer International Publishing, Cham, pp. 72–84
7. Biesialska M, Biesialska K, Costa-jussà MR (2020) Continual lifelong learning in natural language processing:A survey. In Proceedings of the 28th International Conference on Computational Linguistics. International Committee on Computational Linguistics, Barcelona, Spain (Online), pp. 6523–6541
8. Pan JZ, Pavlova S, Li C, Li N, Li Y, Liu J (2018) Content-based fake news detection using knowledge graphs, in International Semantic Web Conference, Springer. pp. 669-683
9. Ghosh S, Shah C (2018) Towards automatic fake news classification. Proceedings of the Association for Information Science and Technology 55:805–807
10. Silva RM, Almeida TA, Yamakami A (2017) MDLText: An efficient and lightweight text classifier. Knowledge-Based Systems 118:152–164

11. Rasool T, Butt WH, Shaukat A, Akram MU (2019) Multi-label fake news detection using multi-layered supervised learning. In Proceedings of the 2019 11th International Conference on Computer and Automation Engineering. ICCAE 2019. Association for Computing Machinery, New York, NY, USA, pp. 73–77
12. Zhou X, V. V, Zafarani R (2020) Fake news early detection: A theory-driven model. Digital Threats: Research and Practice 1 (2)
13. Ahmed H, Traore I, Saad S (2017) Detection of online fake news using n-gram analysis and machine learning techniques, in: International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud EnvKnowledge-Basednger. pp. 127–138
14. O'Brien N, Latessa S, Evangelopoulos G, Boix X (2018) The language of fake news: Opening the black-box of deep learning-based detectors
15. Ruchansky N, Seo S, Liu Y (2017) A hybrid deep model for fake news detection, in Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, ACM. pp. 797–806
16. Roy A, Basak K, Ekbal A, Bhattacharyya P (2018) A deep ensemble framework for fake news detection and classification https://arxiv.org/abs/1811.046
17. Al-Ash HS, Putri MF, Mursanto P, Bustamam A (2019) Ensemble learning approach on Indonesian fake news classification, in Proceedings of the 3rd International Conference on Informatics and Computational Sciences, pp. 29-30, Semarang, Indonesia
18. Ahmad I, Yousaf M, Yousaf S, Ahmad MO (2020) Fake news detection using machine learning ensemble methods, Complexity, 2020(8885861):11
19. Ksieniewicz P, Zyblewski P, Choraś M, Kozik R, Giełczyk A, Woźniak M (2020) Fake news detection from data streams. In 2020 International Joint Conference on Neural Networks (IJCNN). pp. 1–8
20. Das, Bijoyan, Chakraborty, Sarit (2018) An Improved Text Sentiment Classification Model Using TF-IDF and Next Word Negation. arxiv. arXiv:1806.06407
21. Fu X, Liu W, Xu Y, Cui L (2017) Combine hownet lexicon to train phrase recursive autoencoder for sentence-level sentiment analysis. Neurocomputing 241:18–27
22. Cerisara C, Kral P, Lenc L (2018) On the effects of using word2vec representations in neural networks for dialogue act recognition. Comput Speech Language 47:175–193
23. Horne BD, Nørregaard J, Adali S (2019) Robust fake news detection over time and attack. ACM Transactions on Intelligent Systems and Technology 11(1)
24. Mohawesh Rami, Tran Son, Ollington Robert, Shuxiang Xu (2021) Analysis of concept drift in fake reviews detection. Expert Systems with Applications 169. https://doi.org/10.1016/j.eswa.2020.114318
25. Silva RM, Almeida T (2021) How concept drift can impair the classification of fake news. in: symposium on knowledge discovery, mining and learning (kdmile), https://doi.org/10.5753/kdmile.2021.17469
26. Es PAGE (1954) CONTINUOUS INSPECTION SCHEMES. Biometrika 41(1–2):100–115. https://doi.org/10.1093/biomet/41.1-2.100
27. Rohit Kumar Kaliyar, Anurag Goswami, Pratik Narang, and Soumendu Sinha (2020) FNDNet - A deep convolutional neural network for fake news detection. Cogn Syst Res 61, C (Jun 2020), 32-44. https://doi.org/10.1016/j.cogsys.2019.12.005
28. Kaliyar RK, Goswami A, Narang P (2021) FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. Multimed Tools Appl 80:11765–11788. https://doi.org/10.1007/s11042-020-10183-2
29. Allcott H, Gentzkow M (2017) Social media and fake news in the 2016 election. J Econ Perspectives 31(2):211–236
30. Hua J, Shaw R (2020) Corona virus (covid-19) "infodemic" and emerging issues through a data lens: the case of China. Int J Environ Res Public Health 17(7):2309
31. Wong J (2016) Almost all the traffic to fake news sites is from facebook, new data show
32. Holan AD (2016) 2016 Lie of the Year: Fake News. Politifact, Washington, DC, USA
33. Peters M, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, and Zettlemoyer L (2018) Dep contextualized word representations. In: Proceedings of the 2018 conference of the north american chapter of the association for computational linguistics: human language technologies, Volume 1 (Long Papers), pp. 2227–2237
34. Saikia P, Gundale K, Jain A, Jadeja D, Patel H, Roy M (2022) Modelling Social Context for Fake News Detection: A Graph Neural Network Based Approach. arXiv. https://doi.org/10.48550/arXiv.2207.13500
35. A. Bani-Hani, O. Adedugbe, E. Benkhelifa, M. Majdalawieh and F. Al-Obeidat, "A Semantic Model for Context-Based Fake News Detection on Social Media," in 2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA), Antalya, Turkey, 2020 pp. 1-7. https://doi.org/10.1109/AICCSA50499.2020.9316504
36. Es PAGE (1954) CONTINUOUS INSPECTION SCHEMES. Biometrika 41(1–2):100–115. https://doi.org/10.1093/biomet/41.1-2.100

37. Silva M, R., R. Pires P, Almeida TA (2023) Incremental Learning for Fake News Detection. J Inform Data Manage 13(6). https://doi.org/10.5753/jidm.2022.2542
38. Zechao, Li., Jinhui, Tang., Xueming, Wang., Jing, Liu., Hanqing, Lu (2016) Multimedia News Summarization in Search. ACM Transactions on Intelligent Systems and Technology, 7(3):33-. https://doi.org/10.1145/2822907
39. Zechao, Li. (2017) Understanding-Oriented Multimedia News Summarization. 131–153. https://doi.org/10.1007/978-981-10-3689-7-6