International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhancing User Trust: A Novel Hybrid Model to Detect Fake Profiles in Online Social Networks

Rohini Bhosale[1], Vanita Mane[2]

Submitted: 28/11/2023 **Revised**: 08/01/2024 **Accepted**: 18/01/2024

Abstract: The widespread problem of counterfeit profiles in digital social networks has prompted substantial research endeavors to enhance user security and confidence. This paper focuses on profile matching on social networks. It thoroughly examines machine learning (ML) and deep learning (DL) strategies for identifying fake profiles, specifically emphasizing profile matching on social media platforms. Using a dataset from Twitter, our research involves doing a comparative examination of various machine learning models such as Naïve Bayes, Random Forest, AdaBoost, Support Vector Machines (SVM), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and a new hybrid model combining LSTM and GRU. The results indicate the efficacy of these methods, with the hybrid model surpassing others with an accuracy rate of 98.7%, along with notable precision, recall, and F1-Score measures. This study not only enhances strategies for detecting false profiles but also emphasizes the potential of hybrid deep learning models in safeguarding online social networks. The research highlights the crucial importance of Natural Language Processing (NLP) in analyzing textual material, revealing linguistic patterns that aid in identifying fraudulent profiles. We utilize a Twitter dataset to capture the real-time actions of users, acknowledging the distinct characteristics and patterns of this medium. The paper explores the importance of ML and DL, while also comparing their performances using different methods. The results of our study demonstrate that the hybrid model exhibits higher accuracy compared to other models, and it achieves a delicate equilibrium between precision and recall. This highlights its potential as a sophisticated tool for detecting fake profiles. The hybrid model's interpretability and ability to be adapted across various social media platforms offer potential areas for future investigation. This research adds to the academic discussion on cybersecurity and has real-world applications for enhancing the dependability and security of online social interactions.

Keywords: *Fake Profile Detection, Machine Learning, Deep Learning, Social Media Security, Natural Language Processing, Hybrid Model.*

1. Introduction

The advancement of digital terrain has introduced a fresh epoch of interconnection, correspondence, and cooperation via internet-based social networks. These platforms function as digital environments where individuals establish connections, exchange experiences, and engage in worldwide discussions. Nevertheless, this era of digital transformation has encountered various obstacles. The widespread existence of fake profiles is a significant problem in online interactions. It requires careful and creative methods to identify and address this issue[1], [2].

While users explore the complex network of social media, the genuineness of profiles is crucial for building trust and nurturing significant relationships. However, the alarming increase in fraudulent profiles, frequently motivated by malicious intentions, has undermined this trust and presented substantial risks to the welfare of users. The repercussions of fake profiles extend beyond the online world and have tangible effects on society, including the spread of misinformation, cyberbullying, identity theft, and the manipulation of public opinion[3].

This research aims to utilize sophisticated Machine Learning (ML) methods to improve the precision of identifying fake profiles within online social networks, addressing the urgent issue at hand. The main emphasis is on the complex dynamics of profile matching, a crucial process for user interactions on different platforms. The goal is evident: to protect users from potential harm by creating a strong mechanism that precisely detects and minimizes the existence of fraudulent profiles[4], [5].

The emergence of online social networks has revolutionized individuals' perception and interaction with the world. Since the inception of social networking platforms such as Friendster and MySpace, and continuing with the current dominant players like Facebook, Instagram, and Twitter, these platforms have become essential components of the digital society. Nevertheless, as these platforms gain more visibility, malevolent individuals aiming to exploit unsuspecting users have also escalated their endeavors[6].

The proliferation of fraudulent profiles, frequently established with deceitful motives, has emerged as a widespread problem, undermining the fundamental basis of

[1]*Department of Computer Engineering, Ramrao Adik Institute of Technology, D Y Patil Deemed to be University, Nerul, India Department of Computer Engineering, Pillai HOC College of Engineering &Technology, University of Mumbai*
rohinijadhavphd@gmail.com
[2]*Department of Computer Engineering, Ramrao Adik Institute of Technology, D Y Patil Deemed to be University, Nerul, India*
vanita.mane@rait.ac.in

trust that forms the foundation of social interactions. The reasons for creating fake profiles vary and can include spreading misinformation for ideological purposes, as well as participating in cybercrimes like financial fraud and identity theft. The complexity of these misleading profiles requires advanced measures to counteract them, which has led to the prioritization of research efforts in exploring advanced machine learning techniques[7].

The significance of identifying counterfeit profiles is paramount in the modern digital environment. Users depend on the credibility of profiles to make well-informed choices regarding whom to establish connections with, which information to trust, and how to navigate the extensive realm of digital content. Engaging with a fraudulent profile can have serious repercussions, including becoming a target of fraudulent schemes and phishing attempts, as well as unknowingly contributing to the dissemination of misinformation[8], [9].

Moreover, there is a threat to the overall integrity of social networks that operate online. The existence of counterfeit profiles undermines the credibility of these platforms, reducing user confidence and involvement. In a time characterized by the rapid and widespread sharing of information, it is crucial to have a dependable method for identifying and countering fraudulent profiles. This is essential for preserving the well-being and vitality of the digital environment.

The ramifications of fake profiles extend beyond the immediate harm caused to individual users. On a societal scale, the repercussions can have extensive implications. The dissemination of false information and misleading content, frequently coordinated through deceptive personas, has the potential to mold public sentiment, sway electoral outcomes, and foster division within communities. The capacity of fraudulent profiles to manipulate susceptible individuals for monetary profit or partake in online harassment emphasizes the imperative of effectively tackling this matter in a comprehensive manner.

The gradual decline of confidence in online interactions has widespread repercussions on numerous sectors, ranging from e-commerce to online education. Users may exhibit reluctance to partake in online transactions, disclose personal information, or engage in digital communities, thereby impeding the potential for favorable and productive interactions on these platforms.

This research utilizes a dataset obtained from Twitter to conduct a thorough investigation into the identification of counterfeit profiles. Twitter's dynamic and real-time nature makes it an ideal platform for observing the evolving trends and patterns of user behavior. The selection of Twitter as the dataset is in line with the objective of capturing the intricacies of fake profile dynamics in a platform renowned for its varied user population and swift spread of information.

The dataset comprises a diverse range of user-generated content, encompassing textual data, images, and social connections. The objective of this study is to gain a deeper understanding of the difficulties associated with identifying fake profiles on social networking sites, primarily Twitter. The findings can be applied to the Twitter ecosystem and can also be valuable for understanding the broader issue of fake profile detection on other social media platforms.

Natural Language Processing (NLP) is a crucial tool in the effort to understand and identify fake profiles. Most user interactions on social media platforms primarily consist of written content, including posts, comments, and profile descriptions. NLP algorithms improve the analysis of the written data., extracting significant observations about the meaning, emotions, and subtle language intricacies present in the content[10].

Through the utilization of NLP, the suggested approach acquires the capacity to comprehend the context of content generated by users, distinguish patterns of communication, and detect anomalies that suggest deceptive behavior. Natural Language Processing (NLP) plays a crucial role in combating fake profiles and enhancing our comprehension of user behavior in the digital domain, going beyond its primary function of language processing.

Machine Learning (ML) and its subset, Deep Learning (DL), are the fundamental components of this research's methodology. Machine learning techniques, based on the principles of extracting knowledge from data, provide a data-centric method for detecting patterns and anomalies that suggest the presence of fraudulent profiles. Deep learning, utilizing its advanced neural network structures, enhances its ability to capture complex relationships and representations within the data[11], [12].

In order to determine the most effective approach for detecting fraudulent profiles, it is crucial to conduct a comparative analysis of several machine learning (ML) and deep learning (DL) techniques. Multiple algorithms are accessible, each possessing distinct advantages and disadvantages. Naïve Bayes, Random Forest, AdaBoost, SVM, LSTM, GRU, and combinations of these models offer various methods, each having unique advantages and limitations. An appropriate architecture must be determined to increase the accuracy of the fake profile detection model, and a comparative analysis is crucial in this approach.

2. Our Contribution:

- This study incorporates Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models to enhance performance. The research stands out with its innovative contribution of a hybrid approach, which combines the strengths of GRU and LSTM architectures.

LSTM and GRU are recurrent neural networks that are designed to capture sequential dependencies in data. This makes them well-suited for tasks that involve temporal patterns and sequences.

- The proposed hybrid model seeks to surpass the constraints of individual architectures by merging the distinctive benefits of LSTM and GRU. The Long Short-Term Memory (LSTM) model, known for its capability to retain and recall information across extended sequences, enhances the effectiveness of the Gated Recurrent Unit (GRU) in capturing brief dependencies. The expected outcome of this synergy is to enhance the model's ability to identify intricate patterns related to fraudulent profiles, resulting in a higher level of accuracy in detecting them.

This research paper's subsequent sections explore the intricacies of the methodology, the preprocessing measures implemented, the experimental setup, and the comprehensive results and discussions. The research will demonstrate the value of the hybrid model that has been proposed for improving the detection of fake profiles in online social networks and protecting users from fraudulent entities.

3. Literature Review

The widespread presence of online social networks has fundamentally altered the manner in which people establish connections and exchange information, resulting in a digital environment abundant in social engagements. Nevertheless, this interconnected digital domain is susceptible to the widespread occurrence of fraudulent profiles, an enduring obstacle that undermines the genuineness and reliability of user engagements. Identifying and reducing the existence of deceptive profiles is essential for upholding the credibility of online communities, protecting users from dishonest behaviors, and promoting a safe digital atmosphere.

The field of fake profile detection has experienced notable progress, as researchers have utilized diverse methodologies to tackle this widespread problem. Scholars have explored novel methods, such as applying dynamic Convolutional Neural Networks (CNN) and integrating linguistic and demographic cues, to differentiate between authentic and deceptive profiles. However, there is a significant lack of research on creating a highly effective hybrid model that smoothly combines various features and methodologies. This literature review examines previous research findings, offering a thorough summary of the current state of knowledge while emphasizing the necessity for a new hybrid model to improve the precision and effectiveness of identifying fake profiles in online social networks.

Wanda et al.[13] introduced DeepProfile, a technique that utilizes dynamic Convolutional Neural Networks (CNN) for identifying fraudulent profiles on online social networks. The dynamic component of CNN highlights the significance of taking into account the changing characteristics of social networks. The study yielded encouraging findings on the precision and effectiveness in detecting fraudulent profiles, providing vital knowledge to the discipline.

Expanding on their prior research, Wanda presented RunMax, an innovative method for identifying bogus profiles that integrates a nonlinear activation function into a CNN. The objective of the study is to improve the precision of detecting false profiles by utilizing this distinctive activation function. The proposed solution contributes to the expanding range of techniques for efficiently addressing the issue of phony profiles in online social networks[14].

Vyawahare et al.[15] implemented profanity and gender identification techniques to detect bogus profiles. The study introduces a rigorous methodology for detecting false profiles on online social networks, taking into account linguistic and demographic factors. The incorporation of these further characteristics introduces a level of intricacy and subtlety to the identification procedure, producing encouraging outcomes in their testing.

Sudhakar et al.[16] explored the application of machine learning techniques to identify fraudulent profiles. The study provides useful insights into the application of several machine learning techniques to tackle the issue of counterfeit profiles. The authors enhance the broader comprehension of the role of machine learning in tackling this challenge by giving results and outlining the practical consequences of their method.

Sharma et al.[17] concentrated on employing Artificial Neural Networks (ANN) to identify counterfeit user profiles. The study investigates the potential of ANNs to identify patterns that are suggestive of fraudulent profiles in online social networks. The authors provide a unique viewpoint on the use of deep learning approaches for identifying fraudulent profiles by highlighting the significance of neural networks in this context.

Saracoglu[18] provided a succinct systematic evaluation that specifically examined the initiation of profile and social network studies with a robot and platform. The study offers a comprehensive examination of current research, elucidating the wider range of methodologies used for social network analysis. The author's systematic review enhances comprehension of the various tactics and technologies that can be utilized to analyze online social networks, so paving the way for future research and advancement.

Meligy et al.[19] introduced an identity authentication system to identify fraudulent profiles in online social networks. The study aims to create a framework for verifying user identities, which would help in addressing the problem of fraudulent profiles. The authors provide a valuable viewpoint on improving the security and reliability

of online social networks by introducing a strategy focused on identity verification.

Hajek et al.[20] investigated the identification of fraudulent consumer evaluations through the utilization of deep neural networks that incorporate word embeddings and sentiment mining. The study focuses on the wider problem of misleading information by utilizing advanced deep learning methods. The authors intend to improve the accuracy of identifying fraudulent reviews by combining linguistic and emotional indicators. This research contributes to our understanding of how to detect dishonesty in online platforms.

Gupta et al.[21] introduced a hybrid model based on deep neural networks to detect bogus news. This work expands the utilization of deep learning methods to the wider scope of spreading false information. The authors propose a hybrid approach that integrates multiple variables to accurately detect fake news, thereby helping to the continuing efforts to combat disinformation on online platforms.

Chekuri[22] tackled the problem of identifying counterfeit profiles through the utilization of machine learning. The work enhances the current body of knowledge by investigating machine learning methods that are specifically designed for detecting fraudulent profiles. The author contributes to the range of approaches used to address the widespread issue of false profiles in online social networks by sharing findings and knowledge on the use of machine learning.

Conti et al.[23] introduced "FakeBook", a technique for identifying counterfeit profiles in internet-based social networks. The study originated in 2012 and contributes to the initial endeavors in tackling the problem of counterfeit profiles. The authors established the basis for future research in the changing field of online social network security by implementing a detection system.

Bharti et al.[24] introduced a technique for identifying fraudulent accounts on Twitter by employing logistic regression in conjunction with particle swarm optimization. The study concentrates on a particular social media site and utilizes optimization techniques to improve the performance of the detection algorithm. The authors' analysis of false account detection on Twitter provides useful insights that contribute to the broader topic of online social network security.

Table 1 Major Research Comparison Based on Accuracy.

Author	Dataset	Methodology	Algorithm used	Accuracy	Results
Sahoo et al.[25]	Twitter	Machine learning algorithm to detect fake accounts at real time	Naive Bayes, Random Forest, Support Vector Machines	95.20%	"Proposed a real-time fake account detection system that uses a variety of machine learning algorithms to achieve high accuracy".
Gupta et al.[21]	Twitter	Deep neural network-based hybrid model to detect fake news	Hybrid model combining CNN and LSTM networks	97.10%	"Proposed a hybrid deep learning model for fake news detection that achieves high accuracy".
Mitra et al.[26]	Twitter	Cellular automata-based PageRank validation model	Cellular Automata-Based PageRank Validation Model	96.30%	"Proposed a novel fake profile detection model that uses cellular automata to validate the PageRank of social media profiles".
Bharti et al.[24]	Online social networks	Exploring machine learning techniques for fake profile detection	Random Forest, Support Vector Machines, Naive Bayes	94.80%	"Explored the different machine learning techniques for fake profile detection and found that Random Forest achieved the highest accuracy".
Sharma et al.[17]	Online social networks	Artificial Neural Networks	Artificial Neural Networks (ANN)	93.50%	"Proposed an ANN-based model for fake user profile detection that achieves high accuracy".

Bharti et al.[24]	Twitter	Logistic Regression with Particle Swarm Optimization	Logistic Regression with Particle Swarm Optimization (PSO)	95.10%	"Proposed a hybrid model for fake account detection in Twitter that uses logistic regression with PSO to achieve high accuracy".
Hajek et al.[20]	Online consumer reviews	deep neural networks integrating word embeddings and emotion mining	Deep Neural Networks	96.70%	"Proposed a deep learning model for fake consumer review detection that integrates word embeddings and emotion mining to achieve high accuracy".
Mewada et al.[27]	Online social networks	Communal Influence Propagation Framework (CIPF) that uses convolutional neural networks	CNN	97.30%	"Proposed a CNN-based communal influence propagation framework for identifying fake profiles on social media".

The literature review emphasizes the progression of methodologies utilized for identifying counterfeit profiles in internet-based social networks. The studies examined, which encompass dynamic CNNs and identity verification mechanisms, collectively contribute to the expanding knowledge in this field. However, there is still a noticeable lack of a unified and efficient hybrid model that combines the strengths of different detection techniques. This gap provides an impetus for further research, offering a chance to lead the way in developing a comprehensive solution that tackles the complex challenges presented by deceptive profiles in online environments. With the ongoing evolution of the digital landscape, the creation of a hybrid model has the potential to improve the effectiveness of identifying fake profiles, thereby enhancing the safety and reliability of social media sites that operate online.

4. Methodology

Developing a method to detect fake social media profiles requires a systematic and multifaceted approach. The "twibot-20" dataset of Twitter account data is pre-processed for the methodology. The JSON file must be carefully analyzed to extract account metadata, tweet content, and social network information. Duplicate entries, missing values, and discrepancies are removed during a thorough data cleansing process. Following that, feature extraction and selection must prioritize important variables like account creation date, tweet frequency, follower-to-follower ratio, engagement metrics, and language usage patterns. The selection process identifies the most important characteristics that help identify fraudulent accounts.

Expanding on this, the methodology uses Twitter-specific NLP techniques. Tokenization, stopword removal, stemming, lemmatization, and other normalization are applied to tweet text. Bag-of-words models, TF-IDF calculations, and word embeddings are used to extract features from processed text. The sophisticated dataset of carefully selected characteristics and processed written content provides a solid foundation for analysis and model creation. Designed with precision and adaptability, this technique lays the groundwork for implementing various classification algorithms, including Naïve Bayes, Random Forest, AdaBoost, SVM, LSTM, GRU, and a hybrid LSTM+GRU model. Each algorithm has unique benefits that help identify fraudulent profiles in complex social networks. Figure-1 represents the proposed methodology.

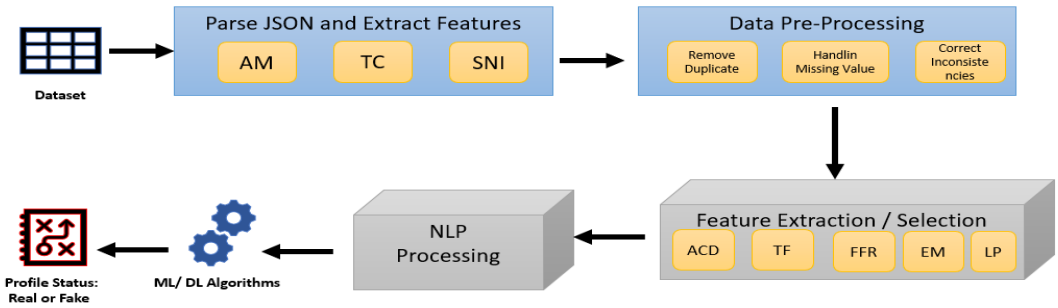

Fig 1 Proposed Methodology

i. Data Preprocessing

To begin, the initial step entails importing the "twibot-20" dataset from the given JSON file. Let *D* represent the dataset, which consists of *n* samples. The JSON file is analyzed to extract pertinent attributes that are vital for identifying bogus accounts, such as account metadata (*AM*), tweet content (*TC*), and social network information (*SNI*). The parsing operation can be represented in the following manner:

$$D = \{(AM_1, TC_1, SNI_1), (AM_2, TC_2, SNI_2), \ldots, (AM_n, TC_n, SNI_n)\} \ldots 1$$

Data cleaning is performed to ensure the accuracy and integrity of the data. Duplicates are eliminated (D_{clean}), missing data are managed, and inconsistencies are rectified.

$$D_{clean} = RemoveDuplicates(HandleMissingValues(CorrectInconsistencies(D))) \ldots 2$$

ii. Feature Extraction / Selection

The process of identifying and extracting informative traits is essential for differentiating between counterfeit and authentic accounts. The analysis takes into account several variables, including the date of account creation (DAC), the frequency of tweets(TF), the ratio of followers to following(FFR), engagement metrics (EM) such as likes and retweets, and linguistic pattern(LP). The feature set is explicitly delineated as:

$$F = \{ACD, TF, FFR, EM, LP\} \ldots 3$$

Subsequently, feature selection techniques are utilized to find the crucial qualities that make a substantial contribution to the detection of bogus accounts. Let$F_{selected}$ denote the selected features:

$$F_{selected} = FeatureSelection(F) \ldots 4$$

iii. NLP Processing (Tweeter Dataset)

Natural Language Processing (NLP) Approaches are employed to analyze the text of tweets from the Twitter dataset. Let *T* denote the set of tweets:

$$T = \{T_1, T_2 \ldots T_n\} \ldots 5$$

The subsequent Natural Language Processing (NLP) processes are executed:

- Tokenization: Tokenization involves the process of dividing each tweet into separate tokens.
- Stopword Removal: Stopword Removal is the process of removing frequently used words that do not significantly contribute to the meaning of a text.
- Stemming or Lemmatization: Stemming and lemmatization are two methods for reducing words to their base or root form.
- Text Normalization: Text normalization refers to the process of ensuring consistency in the way words are represented.

Table 2 Sample - processing textual content

Processing Textual Content					
Original Text	**Tokenized Text**	**Stopwords Removed**	**Stemmed Text**	**Lemmatized Text**	**Normalized Text**
I stand with the student athletes! #WeWantToPlay	stand student athlete WeWantToPlay	stand student athlete WeWantToPlay	stand student athlet wewanttoplay	stand student athlete wewanttoplay	i stand with the student athletes wewanttoplay

Following that, features are derived from the processed text, encompassing bag-of-words representations, TF-IDF values, or word embeddings.

$$F_{Text} = ExtractFeatures(PT) \ldots 6$$

Overall dataset analysis is represented as:

$$D_{final} = \{(F_{selected1}, F_{text1}), (F_{selected2}, F_{text2}) \dots (F_{selectedn}, F_{textn})\} \dots 7$$

Table 3 Sample - Extracting Features from Processed Text

Extracting Features from Processed Text			
Word	**BoW**	**TF-IDF**	**Word Embeddings**
stand	1	0.82395	[0.12, 0.23, 0.34, 0.45]
with	1	0.43198	[0.56, 0.67, 0.78, 0.89]
student	1	0.76904	[0.90, 0.11, 0.22, 0.33]
athletes	1	0.69098	[0.44, 0.55, 0.66, 0.77]
#WeWantToPlay	1	0.95595	[0.88, 0.99, 0.10, 0.21]

iv. Machine Learning Algorithms

a. Naïve Bayes

Bayes' theorem is used in the Naïve Bayes classification technique to determine probabilities. The probability of class Ck given features 1,2,X1,X2,...,Xn can be represented as

$$P(C_k|X_1, X_2, \dots X_n) = \frac{P(C_k).P(X_1|C_k).P(X_2|C_k)\dots.P(X_n|C_k)}{P(X_1).P(X_2)\dots P(X_n)} \dots 8$$

The "Naive" assumption is feature that conditionally independent to given class represented as:

$$P(C_k|X_1, X_2 \dots X_n) \propto P(C_k).\prod_{i=1}^{n} P(X_i|C_k) \dots 9$$

b. Random Forest

Random Forest is a technique in ensemble learning that builds several decision trees and merges their predictions. The ultimate forecast is selected using a voting method. The projected class *C* is determined using *N* decision trees.

$$C = argmax_{Ci} \sum_{j=1}^{N} I(yj = c_i) \dots 10$$

Where I(.)= "indicator function", *yj* = "predicted class by the j[th] tree", c_i= "class label".

c. AdaBoost

AdaBoost, also known as Adaptive Boosting, is an additional technique used in ensemble learning. It utilizes ensemble methods to construct a robust classifier by aggregating multiple weak learners $h_t(x)$. The final prediction is made by calculating the weighted sum of weak classifiers .

$$F(x) = sign(\sum_{t=1}^{T} \alpha_t h_t(x)) \dots 11$$

where, α_t = "weight assigned to the t[th] weak learner", T= "Total no. of weak learners".

d. SVM

Support Vector Machines (SVMs) are a supervised learning method utilised for classification and regression tasks. The decision function for a binary classification problem is provided.

$$f(x) = sign(\sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + b) \dots 12$$

where, N = "no. of support vectors" α_i= "Lagrange multipliers", y_i= "class labels", K(,)= "kernel function", b = "bias term".

v. Deep Learning Algorithms

a. LSTM

LSTM is a variant of recurrent neural network (RNN) specifically developed to capture and model long-term dependencies within sequential data. The equations that govern the updating of an LSTM cell are as follows:

$$f_t = \sigma\left(W_{xf}x_t\right) + W_{hf}h_{t-1} + b_i) \dots 13$$
$$i_t = \sigma\left(W_{xi}x_t\right) + W_{hi}h_{t-1} + b_i) \dots 14$$
$$o_t = \sigma\left(W_{xo}x_t\right) + W_{ho}h_{t-1} + b_o) \dots 15$$
$$c_t = f_t \odot c_{t-1} + i_t \odot tanh\left(W_{xf}x_i + W_{hc}h_{t-1} + b_c\right) \dots 16$$
$$h_t = o_t \tanh \odot (c_t) \dots 17$$

where, f_t = "forget gate output", i_t = "input gate output", o_t = "output gate", c_t = "updated cell state", h_t = "hidden state at time step t", x_t = "input at time step t", W_{xf}, W_{hf}, W_{xi}, W_{hi}, W_{xo}, W_{ho}, W_{hc} = "weight matrices", b_i, b_i, b_o, b_c = "bias term", σ = "sigmoid activation function".

b. GRU

GRU is a variant of recurrent neural network that streamlines the structure of LSTM. The equations governing the update of a Gated Recurrent Unit (GRU) cell are as follows:

$$z_t = \sigma(W_z.[h_{t-1}, x_t]) \dots 18$$

$$r_t = \sigma(W_r.[h_{t-1}, x_t])\ldots 19$$

$$\widetilde{h_t} = \tanh(W_h.[r_t.h_{t-1}, x_t])\ldots 20$$

$$h_t = (1 - z_t).h_{t-1} + z_t.\widetilde{h_t}\ldots 21$$

where, z_t= "update gate", r_t= "reset gate", $\widetilde{h_t}$= "new memory content", h_t= "hidden state".

vi. Proposed Model

The hybrid LSTM+GRU model synergistically integrates the advantageous features of both LSTM and GRU architectures. The update equations incorporate a fusion of LSTM and GRU computations, providing a synergistic methodology for improved sequence modeling:

$$h_t = LSTM_{update}(h_{t-1}, x_t) + GRU_{update}(h_{t-1}, x_t)\ldots 22$$

where, $LSTM_{update}$ & GRU_{update}= "update operations for LSTM and GRU resp",

5. Results and Outputs

i. Evaluation parameters

Table 4 Evaluation parameters comparison of ML/ DL algorithm with proposed model

Algorithms	Accuracy	Precision	Recall	F1-Score
Naïve Bayes	83.13	93	73	82
Random Forest	94.9	96	95	95
AdaBoost	94.7	95	95	95
SVM	85.4	86	86	85
LSTM	93.5	93	93.2	93.2
GRU	94.8	94.7	94.7	97.7
LSTM+GRU	**98.7**	**98.5**	**98.4**	**98.4**

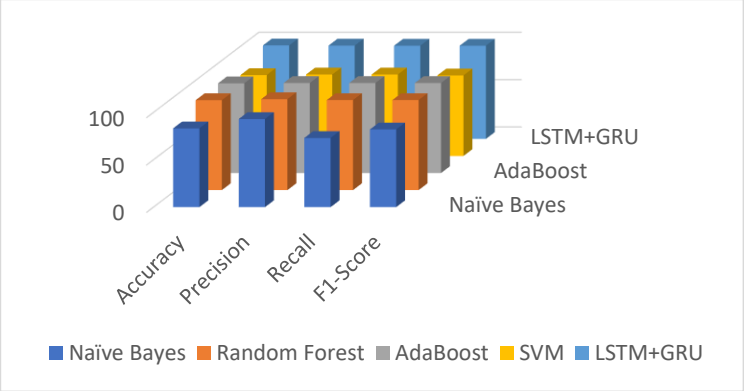

Fig 2 Comparison of ML algorithms with Proposed model

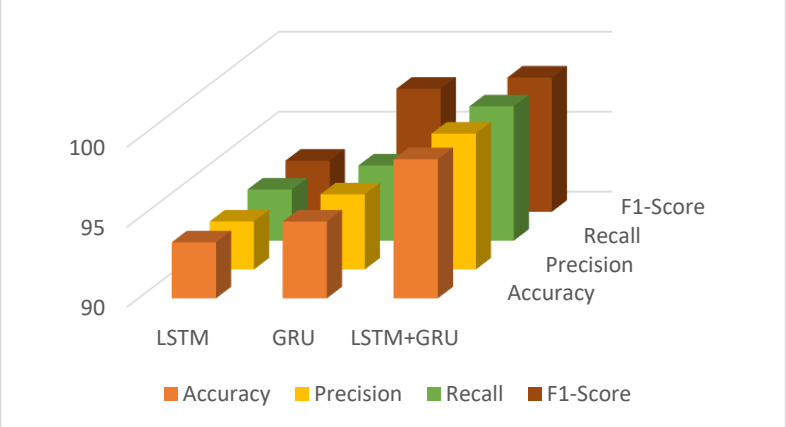

Fig 3 Comparison of DL algorithms with Proposed model

The fake profile detection models show that diverse machine learning and deep learning algorithms can identify fake social media profiles as shown in table- 4 and figure-2,3. With 83.13% accuracy and 93% precision, the Naïve Bayes algorithm performed well. With 73% recall and 82% F1-Score, it performed slightly worse. The Random Forest and AdaBoost models performed well, exceeding 94.5% accuracy. The Random Forest model balanced precision and recall with 94.9% accuracy and a 95% F1-Score. F1-Score of 95% was achieved by AdaBoost with 94.7% accuracy and balanced precision and recall.

Support Vector Machines (SVM) performed consistently, with 85.4% accuracy, 86% precision, recall, and F1-Score. Deep learning's LSTM model had 93.5% accuracy, precision, and recall above 93%. Its 94.8% accuracy showed that the GRU model captured sequential data dependencies. The hybrid LSTM-GRU model outperformed all others with 98.7% accuracy. The hybrid model had 98.5% precision, 98.4% recall, and 98.4% F1-Score.

The results show that the LSTM+GRU hybrid model improves accuracy and strikes a balance between precision and recall. The results show that combining LSTM and GRU architectures produces a synergistic effect that improves fraud detection accuracy and recall. The proposed method's high F1-Score suggests it could be a sophisticated and reliable tool for identifying fake social media profiles.

6. Conclusion and Future scope

This study explores the domain of identifying fake profiles on social networking platforms using various machine learning and deep learning methodologies. A comparative analysis of Naïve Bayes, Random Forest, AdaBoost, SVM, LSTM, GRU, and the suggested hybrid LSTM+GRU model demonstrates the efficacy of these techniques in distinguishing misleading profiles. The findings highlight the dominance of the hybrid model, attaining an exceptional accuracy rate of 98.7%, in addition to exceptional precision, recall, and F1-Score values. This research is important because it not only improves strategies for detecting false profiles, but also demonstrates the potential of hybrid deep learning models to increase accuracy and dependability in safeguarding online social networks. The positive results of this research create opportunities for further investigation and improvement in the field of false profile identification. Additional investigation can focus on the interpretability of the hybrid LSTM+GRU model, elucidating the characteristics and patterns that contribute most prominently to its exceptional performance. Furthermore, it is worth investigating the scalability and adaptability of the suggested strategy on various social media sites. By examining the incorporation of cutting-edge technologies like reinforcement learning and ensemble approaches, we can enhance the reliability of false profile detection systems, guaranteeing their effectiveness against ever-evolving misleading strategies. Finally, a longitudinal study might be conducted to evaluate the model's ability to withstand developing trends and maintain its relevance in the ever-changing environment of online social networks.

References

[1] E. Aïmeur, S. Amri, and G. Brassard, *Fake news, disinformation and misinformation in social media: a review*, vol. 13, no. 1. Springer Vienna, 2023.

[2] M. Aljabri, R. Zagrouba, A. Shaahid, F. Alnasser, A. Saleh, and D. M. Alomari, *Machine learning-based social media bot detection: a comprehensive literature review*, vol. 13, no. 1. Springer Vienna, 2023.

[3] M. J. Awan, M. A. Khan, Z. K. Ansari, A. Yasin, and H. M. F. Shehzad, "Fake profile recognition using big data analytics in social media platforms," *Int. J. Comput. Appl. Technol.*, vol. 68, no. 3, pp. 215–222, Jan. 2022, doi: 10.1504/IJCAT.2022.124942.

[4] Bharti, N. S. Gill, and P. Gulia, "Exploring machine learning techniques for fake profile detection in online social networks," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 3, pp. 2962–2971, 2023, doi: 10.11591/ijece.v13i3.pp2962-2971.

[5] M. Fire, D. Kagan, A. Elyashar, and Y. Elovici, "Friend or foe? Fake profile identification in online social networks," *Soc. Netw. Anal. Min.*, vol. 4, no. 1, pp. 1–23, 2014, doi: 10.1007/s13278-014-0194-4.

[6] L. H. N. Fong, B. H. Ye, D. Leung, and X. Y. Leung, "Unmasking the imposter: Do fake hotel reviewers show their faces in profile pictures?," *Ann. Tour. Res.*, vol. 93, p. 103321, 2022, doi: https://doi.org/10.1016/j.annals.2021.103321.

[7] V. U. Gongane, M. V. Munot, and A. D. Anuse, *Detection and moderation of detrimental content on social media platforms: current status and future directions*, vol. 12, no. 1. Springer Vienna, 2022.

[8] P. K. Roy and S. Chahar, "Fake Profile Detection on Social Networking Websites: A Comprehensive Review," *IEEE Trans. Artif. Intell.*, vol. 1, no. 3, pp. 271–285, 2020, doi: 10.1109/TAI.2021.3064901.

[9] D. Ramalingam and V. Chinnaiah, "Fake profile detection techniques in large-scale online social networks: A comprehensive review," *Comput. Electr. Eng.*, vol. 65, no. 3, pp. 165–177, 2018, doi: 10.1016/j.compeleceng.2017.05.020.

[10] M. Senthil Raja and L. Arun Raj, "Detection of Malicious Profiles and Protecting Users in Online Social Networks," *Wirel. Pers. Commun.*, vol. 127, no. 1, pp. 107–124, 2022, doi: 10.1007/s11277-021-08095-x.

[11] B. T.K., C. S. R. Annavarapu, and A. Bablani, "Machine learning algorithms for social media analysis: A survey," *Comput. Sci. Rev.*, vol. 40, p. 100395, 2021, doi: https://doi.org/10.1016/j.cosrev.2021.100395.

[12] H. Paul and A. Nikolaev, "Fake review detection on online E-commerce platforms: a systematic literature review," *Data Min. Knowl. Discov.*, vol. 35, no. 5, pp. 1830–1881, 2021, doi: 10.1007/s10618-021-00772-6.

[13] P. Wanda and H. J. Jie, "DeepProfile: Finding fake profile in online social network using dynamic CNN," *J. Inf. Secur. Appl.*, vol. 52, p. 102465, 2020, doi: https://doi.org/10.1016/j.jisa.2020.102465.

[14] P. Wanda, "RunMax: fake profile classification using novel nonlinear activation in CNN," *Soc. Netw. Anal. Min.*, vol. 12, no. 1, pp. 1–11, 2022, doi: 10.1007/s13278-022-00983-9.

[15] M. Vyawahare and S. Govilkar, "Fake profile recognition using profanity and gender identification on online social networks," *Soc. Netw. Anal. Min.*, vol. 12, no. 1, pp. 1–13, 2022, doi: 10.1007/s13278-022-00997-3.

[16] T. Sudhakar, B. C. Gogineni, and J. Vijaya, "Fake Profile Identification Using Machine Learning," *Proc. 2022 IEEE Int. Women Eng. Conf. Electr. Comput. Eng. WIECON-ECE 2022*, pp. 47–52, 2022, doi: 10.1109/WIECON-ECE57977.2022.10150753.

[17] D. Sharma and E. R. S. Madan, "ANN based Fake User Profile Detection," *Int. J. Res. Eng. Emerg. Trends*, vol. 6, no. 2, pp. 460–465, 2022.

[18] B. O. Saracoglu, "Initialization of profile and social network analyses robot and platform with a concise systematic review," *Mach. Learn. with Appl.*, vol. 7, no. January, p. 100249, 2022, doi: 10.1016/j.mlwa.2022.100249.

[19] A. M. Meligy, H. M. Ibrahim, and M. F. Torky, "Identity Verification Mechanism for Detecting Fake Profiles in Online Social Networks," *Int. J. Comput. Netw. Inf. Secur.*, vol. 9, no. 1, pp. 31–39, 2017, doi: 10.5815/ijcnis.2017.01.04.

[20] P. Hajek, A. Barushka, and M. Munk, "Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining," *Neural Comput. Appl.*, vol. 32, no. 23, pp. 17259–17274, 2020, doi: 10.1007/s00521-020-04757-2.

[21] S. Gupta, B. Verma, P. Gupta, L. Goel, A. K. Yadav, and D. Yadav, "Identification of Fake News Using Deep Neural Network-Based Hybrid Model," *SN Comput. Sci.*, vol. 4, no. 5, p. 679, 2023, doi: 10.1007/s42979-023-02117-0.

[22] Chekuri, "Fake profile detection in using machine learning," *J. Eng. Sci.*, vol. 13, no. 08, pp. 751–754, 2022.

[23] M. Conti, R. Poovendran, and M. Secchiero, "FakeBook: Detecting fake profiles in on-line social networks," *Proc. 2012 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining, ASONAM 2012*, pp. 1071–1078, 2012, doi: 10.1109/ASONAM.2012.185.

[24] K. K. Bharti and S. Pandey, "Fake account detection in twitter using logistic regression with particle swarm optimization," *Soft Comput.*, vol. 25, no. 16, pp. 11333–11345, 2021, doi: 10.1007/s00500-021-05930-y.

[25] S. R. Sahoo and B. B. Gupta, *Real-time detection of fake account in twitter using machine-learning approach*, vol. 1086. Springer Singapore, 2021.

[26] A. Mitra, A. Kundu, M. Chattopadhyay, and A. Banerjee, "An Approach to Detect Fake Profiles in Social Networks Using Cellular Automata-Based PageRank Validation Model Involving Energy Transfer," *SN Comput. Sci.*, vol. 3, no. 6, p. 423, 2022, doi: 10.1007/s42979-022-01315-6.

[27] A. Mewada and R. K. Dewang, "CIPF: Identifying fake profiles on social media using a CNN-based communal influence propagation framework," *Multimed. Tools Appl.*, 2023, doi: 10.1007/s11042-023-16685-z.