# Subjective Evaluation of Clone Attack Detection Using Machine Learning Approach

Rani, Rupa
Department of Computer Science & Engineering, Banasthali Vidyapith

Kuldeep Kumar Yogi
Department of Computer Science, AIM & ACT, Banasthali Vidyapith

Satya Prakash Yadav
Department of Computer Science and Engineering, G.L. Bajaj Institute of Technology and Management

https://doi.org/10.5109/7236848

出版情報：Evergreen. 11 (3), pp.2011-2021, 2024-09. Interdisciplinary Graduate School of Engineering Sciences, Kyushu University, Japan
バージョン：

# Subjective Evaluation of Clone Attack Detection Using Machine Learning Approach

Rupa Rani[1,2,*], Kuldeep Kumar Yogi[3], Satya Prakash Yadav[4,5]

[1]Department of Computer Science & Engineering, Banasthali Vidyapith, Rajasthan

[2]Department of CSE, Ajay Kumar Garg Engineering College, Ghaziabad, U.P., India

[3]Department of Computer Science, AIM & ACT, Banasthali Vidyapith, Rajasthan

[4]Department of Computer Science and Engineering, G.L. Bajaj Institute of Technology and Management

[5]School of Computer Science Engineering and Technology(SCSET), Bennett University, Greater Noida, U.P., India

*Author to whom correspondence should be addressed:
E-mail: rupachoudhary2010@gmail.com

(Received May 17, 2023: Revised May 10, 2024: Accepted July 21, 2024).

**Abstract**: Social Media Platforms have become an extensive medium for countless people to interact with each other. Social media platforms have become widespread because of the connection among people everywhere in the world and allow sharing of photographs, documents, videos, etc. Privacy and safety challenges among the networks have become a big fear for users generating clone attacks. Here, we are discussing social media platforms where a large number of fake IDs are produced each month and annually, and we briefly explain the research that has been done by the researcher over the past four to five years. We also discuss how to stop these fake IDs and how effective their solutions are. Also, we have summarized the techniques/ algorithms that provide security to social media platforms. Finally, we discussed all the possible solutions to make it better or more secure than before with the help of ML or DL but still, there are not many implementations with advanced machine learning and deep learning algorithms. Research is going on till now. Despite the fact that we cannot entirely secure our social media platforms like Facebook accounts due to the streaming data but cloned attacks can be identified with greater accuracy as compared to previous related research. According to our survey, we can conclude that if we use the latest dataset along with advanced ML and DL algorithms to detect fake accounts and cloned attacks then results can be improved to secure social media platforms.

Keywords: Social Media Platforms; Clone Attacks; Fake Accounts Detection; Machine Learning; Classification Algorithms

## 1. Introduction

A social media platform is a spectacle that is amazing, interesting, or exciting to see that attracts attention. Before this world was not so connected, and talking to people from different beliefs, sharing knowledge, making new acquaintances, sharing photos, videos, and apprenticeships. Social media platforms played a vital title role in globalizing our day-to-day lives. At this moment, as an example in Fig. 1 approximately 2.989 billion[30] people by 2023 (monthly active users) are using at least one Facebook service, unevenly one-third of the world's population mentioned in Fig. 1. People are using social websites such as LinkedIn, Facebook, Twitter, Instagram, and Google+ is on growth. By growing the practice of social websites[8], malevolent operators pursue to disrupt the security of other operators and misuse their terms and authorizations by generating clone accounts, which has become a danger for other operators.

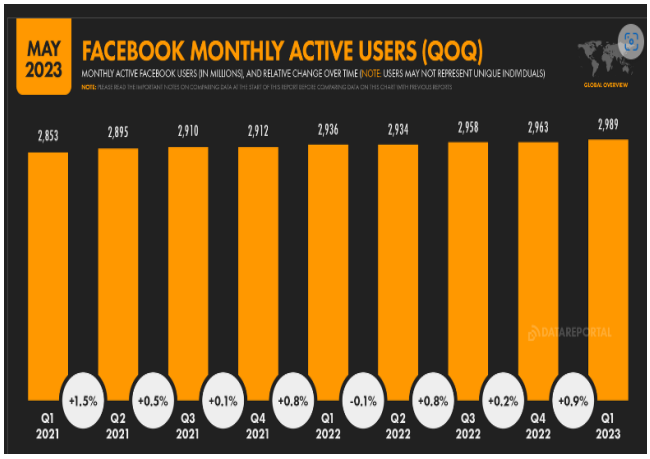


**Fig 1.** Facebook Statistics till Last Quarter of 2022 (2.95 billion) and 2023 (2.989 billion)[30)]

Hence, social media platform providers are trying to find out malevolent operators[9)] and clone accounts to remove them from social websites' surroundings. Generating clone accounts on social media platforms causes more harm as compared to other cybercrime.

social media platforms like Facebook and Big Data[33)] are extending support for mobile safety keys for Facebook iOS/ Android users. Safety keys help ensure that passwords are not the last line of defense between an attacker and your account.

## 2. Literature Survey

Security is a big concern today for everyone, whether they are working or just going about their daily lives. In this section, we have studied several researchers' work and explained below observations obtained by us with the different segmentation and section. The meticulous study is listed as follows.

Mohammadrezaei M.[1)] proposed an approach for detecting clone accounts on social websites, a procedure created on the likeness of the operator's friends. In this procedure firstly friend likeness measures were designed from the adjacency matrix of the system graph and innovative structures were mined from the PCA technique. In the given stage, the information was well-adjusted with the help of SMOTE, and then it was sent to the classifier. This classifier was trained and tested with the help of a cross-validation technique. Here, the Twitter dataset was used for the same. It represented that the Medium Gaussian SVM classifier has a value i.e., AUC (area under the curve) = 1. The user friend network structure was examined in the given procedure and the forgery users were expected by calculating resemblance and the classification algorithms. Here, Medium Gaussian SVM was used in two cases, one for a balanced dataset and the second for an unbalanced dataset in which it provides more accurate results i.e., 97.6% for the balanced dataset as compared to Linear SVM (95.8%) and Logistic regression (96.6%).

Problem is that the fake accounts must perform in the network so that they would be able to identify as clone ones, by exploring friends' networks. This is a drawback of the given method. In the future, a new method can identify the clone account before doing any activity or process of the user in the system or at the time of registration. it can be implemented for more security.

Fire M.[2)] proposed an approach that frightens OSN operators. it can endanger their secrecy, oneness, and good in the actual and pragmatic world both and underlined confident threats that encounter the security of students, youngsters, and teenagers beyond the OSN World Wide Web. Fundamentally, four categories of threats i.e., classic threats, modern threats, combined threats, and threats, targeting children, were explained and solutions were given in three categories i.e., operator solutions, commercial solutions, and category solutions. There is a variety of explanations in table 1 on[2)] that assist to defend an OSN operator's secrecy and safety in this approach but it does not deliver a complete shield to an operator's secrecy and safety. Operators must be attentive to the fact they send online to protect against the variety of online threats and also must apply a variety of solutions. In this, there are eight recommendations (section VI in[2)]) that are very easy to implement for OSN operators to well guard themselves. Moreover, parents must keep watch on their children's pursuits on these social websites, etc. We are obliged to instruct our children to be known of dynamic threats and must communicate to them not to participate with outsiders either in the actual world or in the pragmatic world. Through these social networks, we can improve our survival, but we need to take accurate safeguards to preserve our secrecy and safety.

In the future, a researcher came out with the latest safety and security threats, and evolve estimate and security threats, and evolve and estimate the latest safety outcomes and plans. Universally, investigators can take part in a noteworthy title role by perceiving the use of outcome collaboration and by putting beneficial methods and algorithms. More secure and private algorithms can be implemented at the time of registration.

According to HAYAT M. K.[3)], the social media (SM) program shows a variety of noteworthy questions of deep learning (DL). He described the different-2 SM areas. There are various deep learning techniques that have a noteworthy plan to remember all the things that are of great worth and they can represent the data from different-domain SM programs such as operators' behavior analysis, abnormality identification, sentiment review, business review, etc. Although various views have a strong resource essential to work with the data heaps, improve efficiency, and set down the calculation price, studying well-organized data descriptions from different society information origin and many more, however, need effective and authentic DL-based methods. Furthermore, they will dispatch important changes in several areas of everyday life of a human being like e-commerce, education, medicine, business, and many more.

According to this paper, we have to take high-dimensional data with deep learning algorithms to find an accurate result. It can be implemented in the future.

According to Luceri L.[4], SIDL (social impact in deep learning), is an architecture based on DNN which incorporates deep learning along with network science for modeling and forecasting social supremacy with real-world movements, like catching an affair or going in an emplacement. It performs a pivotal position in forming human behavior & influencing a person's arbitration in innumerable sectors. In this review paper, the proposed work has multiple SIDL techniques which have varied degrees of network association along to overlook two particular issues in deep neural networks i.e., intelligibility and scalability. To resolve the challenge, we instigate G-SIDL, which brings into account the entire social media platform; L-SIDL, considering the ego network of every user; accordingly, C-SIDL, -breaks G-SIDL into tiny sub-replica based on social media associations. It evaluates and validates apply data from propitiating, an occasion-based social media platform, and geometrical, a position-Based Social media platform. Lastly, traverse the usage of various deep learning frameworks, and also provide the observation on the interaction among social impact and people's privacy entrusting consequence and some account of awareness about the possibility of distributing delicate information.

Here G-SIDL that these techniques consult (page no. 20 on[4] ) discloses that these techniques are capable to evaluate a respective behavior with great veracity depending on the insight about friends' events. User action and presence, in turn, can perform, sensitive data, which is a topic that may not want to share in some cases (or future). Though, the current analysis predicts the behavior of humans using the knowledge determined by their friends on an online social media platform. Eventually, we can analyze the passably of using various deep learning frameworks across the architecture by displaying some encouraging impacts and also investigate the relationship between social impacts and users' privacy representing the alarm arrangement.

Aljably R.[5] planned a framework designed to assure the user's protection in OSNs or multimedia social media platforms. ML[35] enamor detects various approaches with ingress standard replica is the strategy impart to the protection of security by notifying users and finding and obstructing irregular actions. This had been done by sending a questionnaire to the user in the dataset to identify the current status of Facebook. It is successfully tested on real data set and got 95% accuracy with the Bayesian classifier & 95.53% accuracy with DNN and LSTM -RNN. Due to too much data on social media e.g., 31.25 million Facebook posts per minute we can use SVM, isolation forest, PCA, etc.

In the future, we can observe our layouts to imply numerous access control architectures and embody other characteristics for ingress assent. We also proposed observation using the adversary drilling method with generative adversarial network (GAN) classifiers to ascertain divergence. The classifier creator maps drilling samples from a clatter precedent to information and implements a comparator that identifies real and dummy data.

Islam M. R.[6] proposed a review paper where MID (misinformation detection) in social media platforms has to acquire an innumerable compact of observation and review a progressive area of the research domain. In this technique, there is the various similar label of deception: false details, allegedly, unsolicited, fake news, and disinformation, and talked about an issue from various points of view on how disinformation misguides the public on social media platforms. Deep Learning is a successful and relevant method to count the disinformation issues on online social media platforms. It can be applied to enhance MID in case of unlabelled and unnecessary data.

Though there are various pieces of work in the form of issues like data clarity, data mass, domain difficulty, accountability[10], explainability, characteristics enhancement, collaborative conclusion, system security, consolidating outstanding information, temporary modeling, etc. which can be enhanced in the next investigation. But there is a forthcoming future scope is to enlarge designing/modeling. So firstly, we can analyze the operator's networks and their previous or old tasks on social media platforms where the operator can think about how they are connected to the open-out forgery news. next, we can integrate the person's mental situation with the operator's previous information, through which we can do a step-ahead analysis of the operator's movement because the propensity to enhance wrong information is correlated with the person's mental situation.

Jain A. K.[7] discussed the various schemes which are linked with the online social media platform viruses (Fig.3) and their responses using dissimilar prototypes, substructures, and encoded procedures that secure the social media framework operators against various kinds of attacks[11]. The attacker can malevolently use linked information (Fig. 2) for unlawful aims. Even there is a lot of danger if attacked kids.

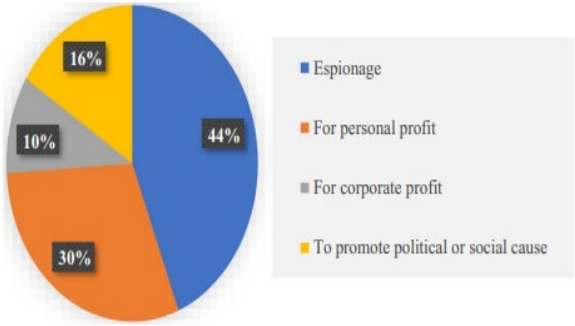


**Fig.2** Punishable Hacking Types in 2021 [7]

Here, the above was explained about the statistics of

social media Platforms. This paper shows the positive and negative factors of social media Platforms [12] . In this, various threats on social media platforms are described in three categories i.e., Conventional, Targeted, and Modern Threats where "spam, identity theft, phishing, etc" comes under Conventional, "Cyber-bullying, Cyber-stalking, etc" "comes in Targeted threats and "Click-jacking, profile cloning, cross-site scripting, etc" comes in modern threat. Such kind of survey describes the open challenges, problems, and related safekeeping instructions to bring out the reliability in online social media platforms.

some issues (Table3 & 4 in[7] ) are necessarily required to be set on by the use of some hybrid techniques, prototypes, and threat identification devices.

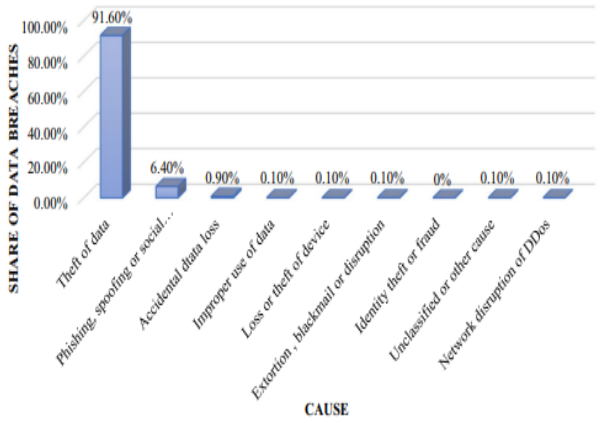


**Fig.3** Worldwide Data Cracks in 2020[7]

Gupta R.[13] proposed an approach where ML-based SDA (secure data analytics) architecture is analyzed with the help of the targeted region, various kind of attacks, and the specifications to reduce such kinds of attacks (table 11 in[13]). Here suggested systems (Fig.4) can find out only what already exists in the database record. It does not have any limit for unusual or hidden attacks because there is no specific description or explanation of various attacks. Here all possible attacks and their vindication methodologies[15] using ML & DL are described. One comparison (Table 1 in[13] ) between the planned and present safe information analytics survey.

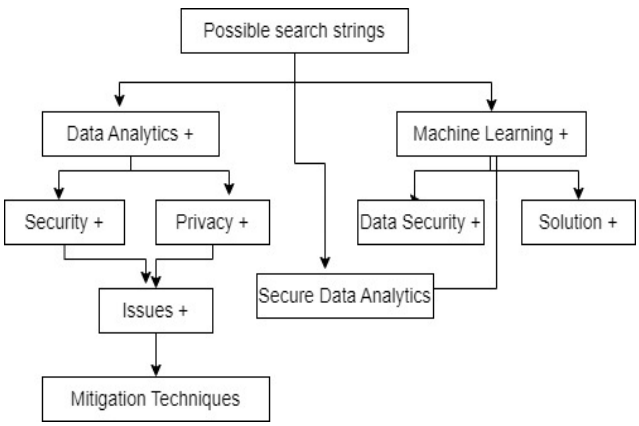


**Fig.4** SNW in Strings[13]

This paper shows the difference between traditional SDA and ML-based SDA architecture. LSTM-RNN can be used in ML-based architecture. It identifies the inappropriate data from input information patterns through learning and recalling the input information pattern which provides safety[17] against the sensor information understanding.

Furthermore, instantaneous-time attacks based on machine learning models would be investigated in the future. Still, there are some research challenges in SDA like data accuracy, multiple attack detection, etc.

Nasir J. A.[14] says that the recent achievement of deep learning procedures is used in complicated natural language processing assignments, constructing an optimistic solution for finding out the forgery news too. Check (table 1 & 2 in[14]). A hybrid model that is a combination of CNN and RNN, is taken for implementation in python on two large actual-world forgery news[19] datasets. Past and present models of ML[18] are used for comparison with this hybrid model. In this architecture, Past trends of the ML model cover supervised learning and ensemble learning, present trends cover un-supervised and semi-supervised learning, and future trends cover reinforcement and deep learning. Through the use of hybrid architecture, we found 100% accuracy on the ISOT dataset (45000 articles) and 60% accuracy on the FA-KES dataset (804 articles).

Task-Specific feature engineering[20] along with complex neural network algorithms can be used in the future for fake news detection.

Kiruthiga. S[21] proposed that a clone account can be detected through the user's click outline and user actions time duration to differentiate between a clone and a genuine one on Facebook. Here the author is using Jaccard similarity[22] and cosine similarity algorithms to improve the performance. The author has taken the Facebook dataset[16] for the experimental analysis. A naïve Bayes classifier is used to identify the network used by any person. K means clustering is used for the same group or school used by the same person. To identify the Cloning attack[23] on Facebook, he is using an enhanced Clone Spotter algorithm (Fig. 5) because of the outline usage and preference list of every user/person. He also prepared a comparison table between Clone Spotter and Enhanced Clone Spotter. Enhanced clone spotter provides more accuracy for clone detection in terms of fake profiles.

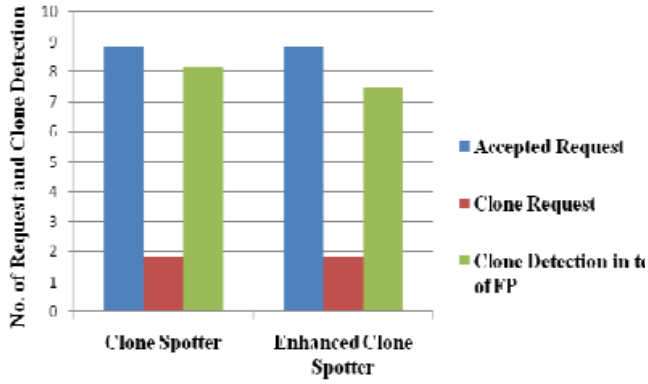


**Fig.5** Clone Attack Detection[21)]

we can use different similarity algorithms to improve the accuracy of clone detection and can also use the latest data set in future.

Khayyambashi M. R.[25)] proposed a methodology to differentiate between genuine and clone profiles by giving an approach of two similarity algorithms along with a predetermined threshold value. First, attribute similarity[24)] is calculated with the help of weighted dice similarity algorithms. Second, Friend network similarity is calculated through three group categorizations. In this, he prepared a friend list, excluded a friend list, and recommended a list graph. This paper checks the profile similarity by using ML and detecting the clone profile {Fig.4, 5 in[21)]). In the future, more accurate results can be obtained through advanced machine learning algorithms.

Agarwal A.[27)] described the performance comparison with the classification ML algorithms. She used the UNSW-NB15 dataset and find the accuracy by using SVM shown in Table.1, Naïve Bayes, and KNN algorithms[26)], earlier it was using the initial BCDR-D01 dataset with the same algorithms. In this, a confusion matrix was generated and some reports are prepared for F1-score, precision, and recall which were used for the training and testing phase of the model. An active and passive intrusion detection system (IDS) was used along with advanced ML algorithms. The latest dataset was minimizing the time of every algorithm.

Table.1 Comparison with the UNSW-NB15 Dataset[27)]

| Serial No. | Classification Algorithms | Accuracy Obtained |
|---|---|---|
| 1 | SVM | 97.7777 (highest) |
| 2 | KNN | 93.3333 |
| 3 | NB | 95.5555 |

It means more datasets provide a more accurate result. Due to the increasing ratio of cyber-attacks, we need to use the future trends of ML-like deep learning algorithms for classification and clustering with the latest dataset so that it could be minimized because it cannot be resolved completely because of social media platforms.

Sophia A.[28)] suggested using Entity Resolution for the detection of fake profiles by using a stenographic algorithm for hiding the facts in pictures. Entity Resolution is a procedure to recognize data records in a solitary data source or transversely several data sources that mention a similar actual-world entity and to connect the record. By using a stenographic algorithm, an id is generated for the profile pictures and we can differentiate the real user and fake user IDs, also an alert message was sent to the real user when a fake one was trying to create the same image (Fig.6). Thus, fake users got blocked. There is no experimental analysis in this paper. In the future, it can be implemented on the latest dataset of Facebook, Twitter, etc.

Kharaji M. Y.[29)] described the identity-cloned attacks in detail. To resolve it, the author created a graph of the social network. According to the similarities, it was divided into different smaller communities then collect all the similar profiles to the real one, and then calculate the strength of the relationship among them through a graph.

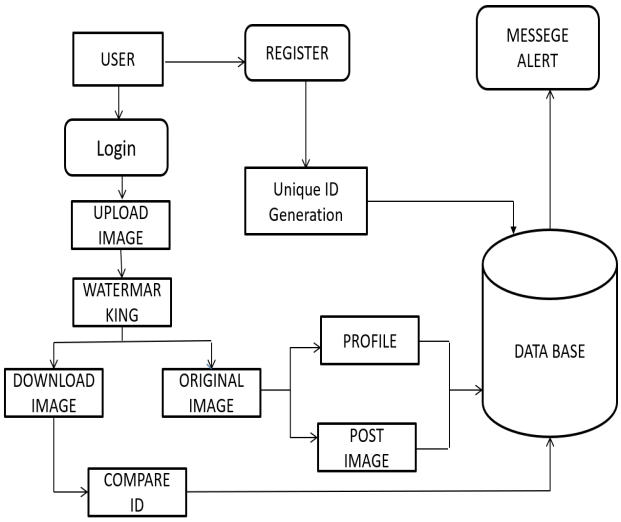


**Fig.6** Architecture to Separate Genuine User ID from Fraudulent ID[28)]

To verify the graph of less strength of relationship, he used the mutual friend system. This work had been done on a Facebook dataset and compared with two previous works. Here IAC clustering algorithm was used to detect the small communities in the public network graph. According to this paper, it is observed that identifying fake identities can minimize the great extent of deception but it is better to prevent than cure so as not to accept the unknown user's request. Also, several fuzzy technologies can be used to overwhelm the incorrectly typed data in the operator's profile in the Facebook application.

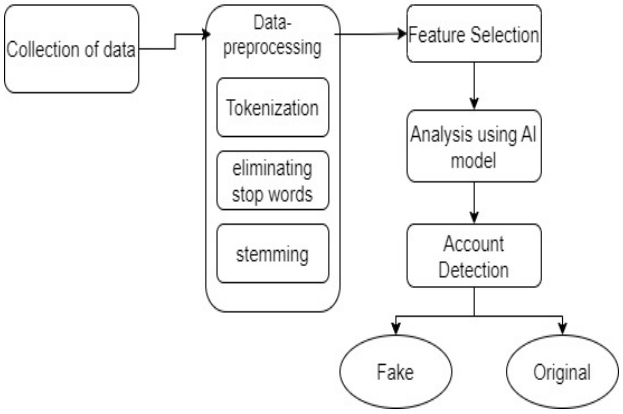


**Fig.7** Identification of Original and Fake Accounts Using AI Model[31)]

Kavin B.P.[31)] explained several spam detection artificial intelligence approaches for Twitter's social media platform. Authors used SVM, ANN, and RF algorithms to identify the performance with user-based and content-based features and authors identifies that SVM is better as compared to others but they used limited types of characteristics and also applied only on Twitter with fewer datasets. In the future, we can use more types of characteristics and can also apply them to other social media platforms with more and the latest dataset.

Saravanan A.[32)] identified the cloned profile by using clustering and classification processes along with big data map reduction techniques. Authors worked on parallel k-means clustering and parallel SVM classification and finally results are verified by using network and attribute similarity. But They used only a sample of 1000 datasets. As a result, Implementation takes less execution time due to map-reduce techniques. They achieved higher accuracy and precision of 98.19 % for the MIB Twitter dataset. In the future, we can work with real-time datasets and also, and we can use any other social media platform that is more in demand.

Table 2. Comparative Study of Previous Related Work

| Authors | Proposed approach | Methods | Performance Parameter | Future scope |
|---|---|---|---|---|
| **Khayyambashi M. R. (2013)[25)]** | Attribute Similarity, friend network similarity algorithms, and weighted graphs | Supervised Learning algorithms | only check profile similarity and find out the clone profile | accurate results can be obtained through advanced machine learning algorithms |
| **Kiruthiga. S (2014)[21)]** | K means clustering and Naïve Bayes Classifier algorithms | Machine Learning Algorithms | Clone accounts detection rate is high because of using enhance clone spotter algorithm. | performance measures can be increased by using the latest ML algorithms. |
| **Mohammadreza (2018)[1)]** | logistic regression, linear Support Vector Machine (SVM) algorithm, Medium Gaussian SVM | Machine Learning | 97.6 % accuracy (highest) by using Medium Gaussian SVM | Deep Learning Algorithms and feature engineering can be used for more accuracy. |
| **Sophia A. (2020)[28)]** | Stenographic algorithm | Cryptography techniques | Fake users can be blocked by differentiating between real and Fake's id through an alert message | It can be implemented on the latest dataset of Facebook, Twitter, etc |
| **Aljably R. (2020)[5)]** | ML anomaly detection techniques | Machine Learning and Deep Learning | 95% accuracy with Bayesian classifier on a real data set and 95.53% accuracy with DNN and LSTM -RNN | Adversarial training procedure with the generative adversarial network can be implemented to detect abnormality |
| **Gupta R. (2020)[13)]** | ML-based SDA architecture (unsupervised and semi-supervised learning algorithms) | Machine Learning & Deep Learning | Provide more accuracy as compared to traditional SDA architecture | Furthermore, actual-time clone attacks on the ML techniques would be investigated. |
| **Nasir J. A. (2021)[14)]** | hybrid approach (RNN and CNN) for forgery news classification | Deep Learning | more efficient and accurate model on a specified dataset | More complex neural network & Task-specific feature engineering can be implemented in the future. |
| **Agarwal A. (2021)[27)]** | Worked on the Limited dataset with SVM, NB, and KNN and no. of samples are very less. | Machine Learning | SVM- 97.777<br>KNN-93.333<br>NB-95.5555<br>The highest accuracy was obtained with the latest dataset | Deep Belief Network, LSTM, and Random Forest can be used with the latest dataset to find more accuracy. |

| | | | | |
|---|---|---|---|---|
| **Kavin B.P. (2022)[31)]** | Applied on the limited dataset and only on Twitter with SVM, RF, and ANN | Machine Learning | SVM is better as compared to others but performance is identified with user-based and content-based features | Social networking website that is more popular among all social media platforms, can use more parameters and the latest big dataset with advanced ML and DL algorithms to identify exact output. |
| **Saravanan A. (2023)[32)]** | Worked on only a sample of 1000 datasets of Twitter. Parallel k-means clustering and parallel SVM classification algo are used with big data map-reduce techniques. | Machine Learning and Big Data map-reduce techniques | Implementation takes less execution time so this model is efficient. | A real-time large dataset is needed to achieve higher and more accurate results. |

## 3. State of Art of the Studied

The number of social media platforms like Facebook users who have signed up to date is 2.963 billion and 1.5 billion FB user data are available for sale online. through which hacking, spamming, scamming, interception, etc are possible among which 6 million users' information is available online from India. There are 533 million (see Fig. 8) FB users' account data (approximately 20% of all accounts) leaked online. Also, it is difficult to manage such kind of big data due to streaming data like during/after Covid-19[34)], data is increased at a very large scale. Because of this, it is very difficult to control the cloned attacks with full accuracy.

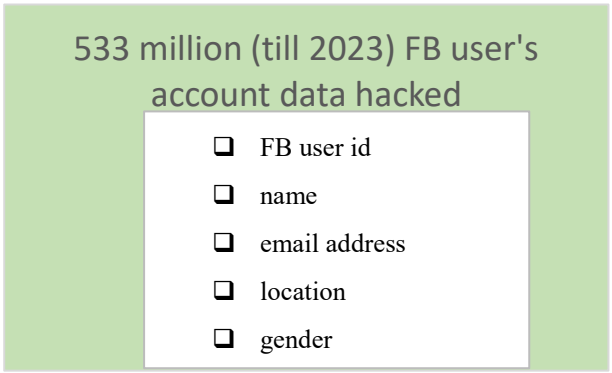


**Fig.8** Basic Information to Hack the Data[30)]

It is concluded in Table. 2 that we must be on the safe side to take care of our security & safety. Our account information can be leaked through a number of different factors, including FB user id, name, email address, location, gender, age, phone number, public profile, etc.

Thus, all of the above represent that there is no option in the "security & privacy setting" [ see the table.3] to identify the clone attacks on Facebook i.e., a social media platform, and only detection methodologies with different accuracies are available in most of the previous researches.

Table.3: Various Threats along with Solutions in the Social Media Platform

| Threats / Solution | Cyber Bulleying | Online Predators | Identify Clone Attacks | Fake Profiles | Face Recognition | Phishing Attack | Spammers |
|---|---|---|---|---|---|---|---|
| Authentication Mechanism | × | ✓ | ✓ | ✓ | × | ✓ | ✓ |
| Security & Privacy Setting | ✓ | ✓ | × | ✓ | ✓ | × | ✓ |
| Fake Profile Detection | ✓ | × | ✓ | ✓ | × | × | ✓ |
| Coned Profile Detection | × | × | ✓ | ✓ | ✓ | × | × |

In previous research, researchers are using different ML algorithms in which most of the experimental analyses are based on past trends and present trends algorithms of ML with different-2 datasets of Facebook, Twitter, etc. In my opinion, there is a very limited number of experimental analyses with the future trends algorithms of ML which covers Deep Learning (CNN and RNN) and reinforcement learning. Only Nasir J.A. [14)] used the concept of deep learning for experimental analysis with a specified dataset. Also, Researchers can improve the recent ultra-model of security products or can enhance security before any activity.

## 4. Assessment of Review

In order to identify fake accounts and cloned attacks on social media platforms, we are employing the RBM_CNN Fusion method. According to this approach, I have taken a victim named Rupa and a cloned Rupa and sent the request to Rupa's family, relatives, and friends. Also, there are two more cloned Rupa's family, relatives, and cloned Rupa's friends, they both also send the request to Rupa's family, relatives, and friends. If anyone accepts the request then information is visible to attackers, and they will also get cloned. Now attackers collect the information about

Rupa and send the request to Rupa and her friends. To eliminate such kinds of issues or cloned attacks, we are using two deep learning methods RBM (Restricted Boltzmann Machines) and CNN (Convolutional Neural Network), and one more RBM_CNN Fusion Approach. Also, there is a comparison among all these three to measure the performance.

The following are the objectives of our research:

- To access the Social Dilemma dataset as input
- To identify the predominant parameters for detecting fake accounts effectively and optimistically.
- To train the deep learning model by bringing out the latest technology-based DL methods and algorithms i.e., the RBM_CNN Fusion approach with the help of the latest dataset.
- To develop a DL-based method (combination of RBM and CNN) on a benchmark database.
- To compare the result of the RBM_CNN Fusion approach (and also to analyze the data visualization through the pictorial view) with the result of previous research like SVM, Naïve Bayes, CNN, etc.

Table 4. Gap Analysis

| Authors | Year | Research Gap |
|---|---|---|
| **Mohammadreza[1)]** | 2018 | It can recognize the clone accounts but ids cannot be detected because of less datasets. |
| **HAYAT M.K.[3)]** | 2019 | It cannot provide much effective and reliable data due to more and more SMA problems |
| **Islam M.R.[6)]** | 2020 | False information and forgery news are still happening at a very large level. It's an active topic and still, and research is going on. |
| **Nasir J.A.[14)]** | 2021 | It can be used only for specified data sets for forgery news classification because of the more complex neural network. |
| **Kavin B.P.[31)]** | 2022 | Here, performance is identified with user-based and content-based features only. SVM is better but only with a limited dataset and also implemented only on Twitter i.e., one of the social media platforms. |
| **Saravanan A.[32)]** | 2023 | Here ML algorithm and big data map-reduce techniques are used for fast execution but it was not implemented on real-time datasets. |

## 5. Architecture of Proposed Approach

The connecting media heavily utilizes a social media platform, which is what we'll employ. Following is the architecture in Fig.9 of the RBM_CNN Fusion approach in which there are two types of filtering one is collaborative filtering in RBM and the second one is filtering in CNN. Through collaborative filtering, we are extracting the features as output and that output will work as an input for CNN to enhance the collaborative filtering. Now we will get the final output either a fake or real account and the information goes to the admin. After that admin will send an alert message to the user if there is any fake account detected. Also, the admin will block such kinds of identities permanently or will represent them as a message i.e., suspicious identity.

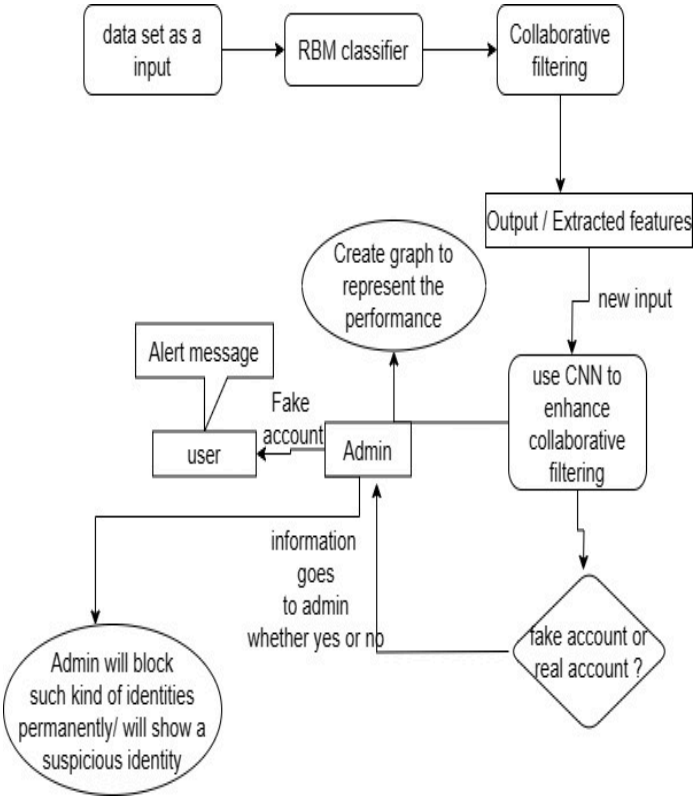


**Fig. 9** Architecture of RBM_CNN Fusion Approach

The proposed system includes the following methodology:

- Data Extraction i.e., Facebook Latest Dataset (Kaggle.com)
- Data pre-processes through one-hot encoding
- Data Partition: Training dataset (70%) and Testing Dataset (30%)
- Building RBM_CNN Fusion Model which consists of two input layers, multiple hidden layers, and 1 output layer with activation function ReLU at the hidden layer and Softmax at the output layer.
- Training and Testing datasets are applied to RBM_CNN Fusion Model
- Evaluating for accuracy
- Compare the result of the "RBM_CNN Fusion

Approach" with the previous latest research-> best result means high accuracy.

## 6. Result

Our survey result shows that we can have two options to improve security or for the detection of cloned attacks and fake profiles. First, we can provide security at the time of registration on any social media platform. At present, we are using a pin or a password for signup to the social media platform but we can have a real-time face recognition password at the time of registration. Second, we can detect and prevent social media platforms to identify cloned attacks and fake accounts by improving the accuracy of the result as compared to previous related research. Day-by-day data is increasing so we cannot provide 100% security but by using big data map-reduce techniques and deep learning algorithms, we can enhance security.

The result of our proposed approach will enhance security by detecting cloned attacks and fake accounts using the RBM_CNN approach i.e., a fusion approach. Also, it is a combination of two deep learning algorithms which is very suitable and best with the latest big dataset.

## 7. Conclusion

Clone accounts can be identified using a variety of machine learning (ML) or deep learning (DL) based approaches [36] or models, according to prior or previous years' research on social media platforms. There are various experimental analyses through ML. Some of them are not using real-time datasets and have higher accuracy with limited datasets, some are using the latest dataset but these have less accuracy than the previous implementation. According to my survey, social media platforms can still be hacked on a large scale. Researchers tried to resolve the same through third-party verification, but the public still faces the problem of hacking the accounts. Here, we are using real-time data sets and the Fusion approach which is a combination of advanced ML and DL algorithms i.e., the RBM_CNN Fusion Model through which we are finding an efficient and high-accuracy result. There is no previous related research available with the same approach till now. however, we cannot find fully accurate results because of the streaming data on social media platforms.

## 8. Future Work

In the future, it will generate additional vigorous and effective security outcomes for finding out the forgery profiles, clone identity attacks, fishing attacks, spammers, and other threats, etc. due to the increasing number of cyber threats. Also, Researchers can improve the recent ultra-model of security products or can enhance security before any activity in social media platforms.

Several people's attributes like the Resemblance between Facebook Liking and other digital information like browsing histories, exploration inquiries, or buying histories advise that the ability to show the user's specific is unlikely to be limited to preferences. Additionally, several attributes were predicted that contain appropriate training data, it may be probable to tell other attributes as well.

## References


1) Mohammadrezaei, M., Shiri, M., E., and Rahmani, A., M. "Identifying Fake Accounts on Social Networks Based on Graph Analysis and Classification Algorithms", Wiley, Volume 2018, Article ID 5923156, 8 pages, August 2018
2) Fire, M., Goldschmidt, R., and Elovici, Y., "Online Social Networks: Threats and Solutions", IEEE COMMUNICATION SURVEYS & TUTORIALS, VOL. 16, NO. 4, 2019
3) HAYAT, M., K., DAUD2, A., ALSHDADI, A., A., BANJAR, A., ABBASI, R., A., BA, Y., AND DAWOOD, H. "Towards Deep Learning Prospects: Insights for Social Media Analytics" IEEE ACCESS, April 3, 2019.
4) Luceri, L., Braun, T., and Giordano, S. "Analysing and inferring human real-life behaviour through online social networks with social influence deep learning", Springer, 25 pages, 2019
5) Aljably, R., Tian, Y., and Rodhaan, M., A. "Preserving Privacy in Multimedia Social Networks Using Machine Learning Anomaly Detection", Hindawi (Security and Communication Networks) Volume 2020, Article ID 5874935, 14 pages, 20 July 2020
6) Islam, M., R., Liu, S., Wang, X., Xu, G., "Deep learning for misinformation detection on online social networks: a survey and new perspectives", Springer (Social Network Analysis and Mining), September 2020
7) Jain, A., K., Sahoo, S., R. and Kaubiyal, J. "Online social networks security and privacy: comprehensive review and analysis", published 1st June 2021, Springer. https://doi.org/10.1007/s40747-021-00409-7
8) Pandey B, Bhanodia PK, Khamparia A, Pandey DK (2019) A comprehensive survey of edge prediction in social networks: techniques, parameters and challenges. Expert Syst Appl 124:164–181. https://doi.org/10.1016/j.eswa.2019.01.040
9) Peng S, Zhou Y, Cao L, Yu S, Niu J, Jia W (2018), Influence analysis in social networks: a survey. J Netw Comput Appl 106:17–32. https://doi.org/10.1016/j.jnca.2018.01.005
10) Sarmah U, Bhattacharyya DK, Kalita JK (2018) A survey of detection methods for XSS attacks. J Netw Comput Appl 118:113–143. https://doi.org/10.1016/j.jnca.2018.06.004
11) leszka M (2018) Application of collective knowledge

diffusion in a social network environment. Enterp Inf System 1–23
12) Tse YK, Loh H, Ding J, Zhang M (2018) An investigation of social media data during a product recall scandal. Enterp Inf System 12(6):733–751. https://doi.org/10.1080/17517575.2018.1455110
13) Gupta, R., Tanwar, S., Tyagi, S., Kumar, N. "Machine Learning Models for Secure Data Analytics: A taxonomy and threat model", Elsevier, volume 153, 1 March 2020, Pages 406-440
14) Nasir, J., A., Khanb, O., S., Varlamis, I. "Fake news detection: A hybrid CNN-RNN based deep learning approach", volume 1, issue 1, April 2021, Elsevier. https://doi.org/10.1016/j.jjimei.2020.100007
15) M. Kosinski, D. Stillwell, and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," Proc. Nat. Acad. Sci. USA, vol. 110, no. 15, pp. 5802–5805, Apr. 2013 https://doi.org/10.1073/pnas.1218772110
16) L. Popov, Staying in Control of Your Facebook Logins, May 2010. [Online]. Available: https://www.facebook.com/blog/blog.php?post=389991097130
17) Senthil Kumar N*, Saravanakumar K, Deepa K, "On Privacy and Security in Social Media – A Comprehensive Study", International Conference on Information Security & Privacy (ICISP2015), 11-12 December 2015, Elsevier
18) Rathorea, S., Sharma ,P., K., Loia, Young-Sik Jeongc , Jong Hyuk Parka,∗,"Social network security: Issues, challenges, threats, and solutions", August 2017, Elsevier http://dx.doi.org/10.1016/j.ins.2017.08.063
19) G. Kontaxis, I. Polakis, S. Ioannidis, and E. Markatos, "Detecting social network profile cloning," in Proc. IEEE Int. Conf. PERCOM Workshops, 2011, pp. 295–300.
20) L. A. Cutillo, R. Molva, and T. Strufe, "Safebook: A privacy-preserving online social network leveraging on real-life trust," IEEE Commun. Mag., vol. 47, no. 12, pp. 94–101, Dec. 2009.
21) Kiruthiga. S, Kola Sujatha. P, Kannan. A, "Detecting Cloning Attack in Social Networks Using Classification and Clustering Techniques" International Conference on Recent Trends in Information Technology, 2014, IEEE
22) Rathore, S., Sharma, P., K., Loia V., Jeong, Y., S., Park, J., H. "Social network security: Issues, challenges, threats, and solutions", Volume 421, December 2017, Pages 43-69, Elsevier
23) Sharma, A., Dhawan, S., Sing, K. "A REVIEW PAPER ON PROFILE CLONE DETECTION IN SOCIAL NETWORKS", INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY, [Sharma* et al., 5(7): July, 2016]
24) Rizi, F., S., Khayyambashi, M. R., and Kharaji, M. Y. "https://arxiv.org/ftp/arxiv/papers/1406/1406.7377.pdf A New Approach for Finding Cloned Profiles in Online Social Networks", Int. J. of Network Security, Vol. 6, April 2014
25) Khayyambashi, M., R., Rizi, F., S., An approach for detecting profile cloning in online social networks, 7th International Conference on e-Commerce in Developing Countries with Focus on e-Security, IEEE, 17-18 April, 2013
26) Aashrith, B., A., Ajay, K., D., Anil, K., D., N., Mohammed, M., R., A., Ramesh, G. "Colluding Identity Clone Prediction using Machine Learning", Vol. 8, Issue 4, April 2019, International Journal of Advanced Research in Computer and Communication Engineering
27) Agarwal, A., Sharma, P., Alshehri, M., Ahmed, A., M. and Alfrraj, O. "Classification model for accuracy and intrusion detection using machine learning approach", 7 April, 2021, PeerJ Computer Science
28) Sophia, A., Mukilan, Prasath, G. "RECOGNIZINGUSERPORTRAITFOR FRAUDULENTIDENTIFICATION ON ONLINEAPPLICATIONS", International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 03 | Mar 2020
29) Kharaji, M., Y. and Rizi, F., S. "An IAC Approach for Detecting Profile Cloning in Online Social Networks", International Journal of Network Security & Its Applications (IJNSA), Vol.6, No.1, January 2014
30) Chart: Has Facebook Reached Its Growth Limit? | Statista Or The Latest Facebook Statistics: Everything You Need to Know — DataReportal – Global Digital Insights (2023)
32) Kavin, B., P., Sagar, K., "Machine Learning-Based Secure Data Acquisition for Fake Accounts Detection in Future Mobile Communication Networks", Hindawi, Wireless Communications and Mobile Computing, Volume 2022, Article ID 6356152, 10 pages. https://doi.org/10.1155/2022/6356152
32) Saravanan, A., Venugopal, V. "Detection and Verification of Cloned Profiles in Online Social Networks Using MapReduce Based Clustering and Classification" International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING, ISSN:2147-67992, January 2023.
33) Khan, S., A., Arora, R., Kumar, H., Arora, P., K. " A Perspective on Advances in Cloud-based Additive Manufacturing" EVERGREEN Joint Journal of Novel Carbon Resource Sciences & Green Asia Strategy, Vol. 09(3), pp861-869, (2022). https://doi.org/10.5109/4843119
34) Bansal, M. , Agarwal, A., Pant, M., Kumar, H. "Challenges and Opportunities in Energy Transformation during COVID-19", EVERGREEN

Joint Journal of Novel Carbon Resource Sciences & Green Asia Strategy, Vol. 08(2), pp255-261, (2021). https://doi.org/10.5109/4480701

35) Singh D., Singh, A. "Role of Building Automation Technology in Creating a Smart and Sustainable Built Environment", EVERGREEN Joint Journal of Novel Carbon Resource Sciences & Green Asia Strategy, Vol. 10(1), pp412-420, (2023). https://doi.org/10.5109/6781101

36) R. Rani, K. K. Yogi, and S. P. Yadav, “Detection of Cloned Attacks in Connecting Media using Bernoulli RBM_RF Classifier (BRRC),” *Multimed Tools Appl*, Feb. 2024, doi: 10.1007/s11042-024-18650-w.