




A novel machine learning and face recognition technique for fake accounts detection system on cyber social networks

Ala Mughaid¹ · Ibrahim Obeidat¹ · Shadi AlZu'bi² · Esraa Abu Elsoud¹ · Asma Alnajjar¹ · Anas Ratib Alsoud³ · Laith Abualigah^{4,5,3,6,7,8} 

Received: 15 June 2022 / Revised: 14 November 2022 / Accepted: 2 January 2023 /

Published online: 31 January 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Online Social Networks (OSN) such as Facebook, Instagram, Twitter, and others have seen rapid growth in recent years. Such applications provide attractive online social networks and communications with the opportunity to connect with relatives and acquaintances, meet new people, enter communities, talk, exchange photos, organize events, and network with others who are close to real-life; unfortunately, on the other hand, they also raise privacy and security issues. We identified OSN threats in this paper and recommended a digital face-processing authentication method as a double-factor authentication after entering the password using Matlab. After applying deep learning classification by attending to a real dataset from the live webcam to train the model, we achieved the best accuracy rate of 95%. However, such methods have yet to be deployed to all social networks, so we also mentioned the problem of fake accounts, which is one of the most significant problems in OSN. These are effective tools for executing spam campaigns and spreading malware and phishing attacks. Fake accounts could lead to the loss of money for businesses, loss of reputation, stealing information for malicious purposes, and much more. This study is related to detecting fake and legitimate profiles on OSN. For this purpose, we chose two datasets that contain fake and legitimate accounts on Facebook and Instagram. Each contains different features after applying machine learning using Naive Bayes, Logistic Regression, Support Vector Machines, K-Nearest Neighbour, Boosted Tree, Neural Networks, SVM Kernel, and Logistic Regression Kernel. SVM achieved the highest classification accuracy for the Fake Profiles detection datasets with 97.1%.

Keywords Social media · Cyber attacks · Social media privacy · Computer vision in social media · Social media security · Face recognition

✉ Laith Abualigah
Aligah.2020@gmail.com

1 Introduction

Social media usage is one of the most popular online activities [2]. The number of social media users is increasing every day [23]. For instance, Facebook, WhatsApp, and TikTok are now the preferred choices for many people, as their users spend a generous amount of time on these applications [4, 28, 30]. In 2020, according to Statista.com, the number of people using social media in the world was 3.6 billion, which is expected to increase in 2025. Over 3.6 billion people were using social media worldwide, according to the global platform Statista.com, a number expected to increase to almost 4.41 billion in 2025 [13, 28, 54].

The global social media usage rate was 49 percent in January 2020. Most of the global growth of social media is driven by the growing use of mobile devices. East Asia's mobile market ranks first in the world's mobile social penetration rankings, followed by established digital powers such as America and Northern Europe [21]. The first social network regarding the number of users and registered accounts is Facebook, with more than one billion accounts, with 2.89 billion active users each month. Figure 1 represents the most popular social networks worldwide until October 2021.

Users have become much more prone to sharing their ideas, feelings, and experiments with friends and also with friends of friends through videos, pictures, photos, etc., and this data is publicly available, so preserving privacy in publishing social network data becomes a paramount concern, especially with the explosive growth of cyber security incidents on social networks [3, 64]. These privacy concerns become more alarming when considering the nature of OSNs, which share information that the user can access even without directly accessing the individual's online profile, where the attacker can collect data about the user through the user's friend [1, 45].

Recently, users and their details have been the primary target of many online attacks, including identity theft, inference attacks, malware, click-jacking, phishing attacks, Sybil attacks, reverse social engineering, and social bots. Attackers attempt to collect users' personal information and gain their trust by adding them through fake spam accounts [52, 57]. By using this personal information, an attacker can send personal spam messages in an attempt to redirect them to malicious websites [37].

The attackers attempt to cover their tracks by using fake profiles. The millions can count the number of fake profiles in OSN. For example, Facebook estimates that around 5% of the users could be fake or duplicate accounts [61]. Many researchers have been working on securing data on social media platforms, such as in [5, 7, 25, 26, 51]. OSN operators have offered multiple solutions to handle these threats, such as adding authentication processes to ensure that the registered user is the real one. Also, machine learning classification gives us knowledge about the security levels we need to maintain in social networking, as it proved significant improvements in many human applications [9, 40, 47].

In this paper, we analyzed the online social networks (OSNs) threats and recommended a digital face processing authentication method using MATLAB in the simulation process. We applied advanced machine learning techniques to present a detection model of spam and legitimate profiles in social networks using two datasets with different features to cover up all possible features that could detect fake accounts rather than legitimate accounts using MATLAB. Experiments are conducted, and the results are evaluated using standard measures to prove the performance of the proposed method. The results showed that the proposed method is efficient in dealing with threats. The main contributions of this paper are given as follows.

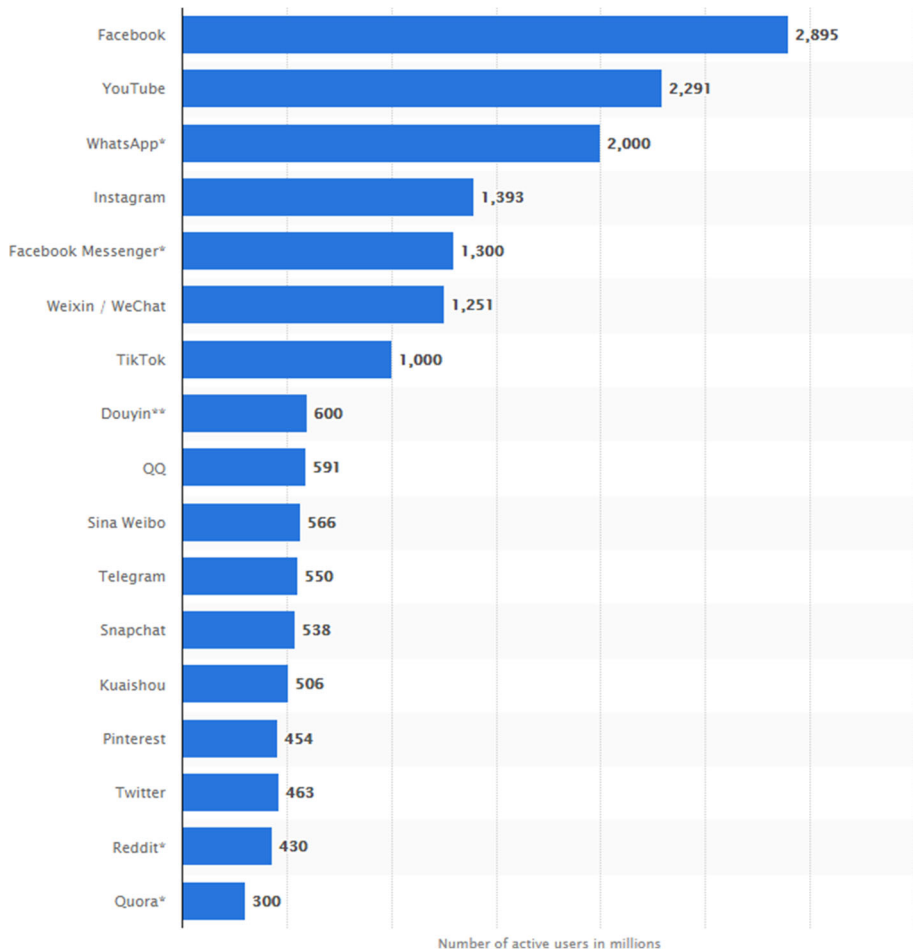


Fig. 1 Number of active users in millions

- The online social networks (OSNs) threats are analyzed using the digital face processing authentication method.
- Advanced machine learning techniques are applied to present a detection model of spam and legitimate profiles in social networks.
- Two datasets with different features are used to cover up all possible features that could detect fake accounts rather than legitimate accounts.
- The results show that the proposed method is efficient in dealing with threats.

The remainder of the paper is organized as follows. In Section 2, we review the related work. Next, Section 3 describes social network types, architecture, threats, and solutions. In Section 4, we discussed our methodology. Finally, the conclusions are presented in Section 5.

2 Literature review

Artificial intelligence has been employed with many human applications such as intelligent transportation systems [27, 29], smart medicine [8, 16, 35], intelligent health [12], smart agriculture [15, 31], renewable energy [11], as well as intelligent business applications [10, 14, 48] and many other applications including social media platforms [46]. The vast success and widespread that online social networks (OSNs) are experiencing nowadays are widely evident. Users use online social networks as a communication channel. OSNs are critical applications because users and their resources are secure. They are attracting more malicious attacks that, concerning traditional ones, have the advantage of being more effective by leveraging the social network as a new medium for reaching victims.

Iliia et al. in [41] designed an access control mechanism that allows users to set their preferred permissions based on their face. Their approach is designed for easy integration with existing social networks. Al Hasib in [6] studied the most critical threats to social network services and presented technical solutions to improve security and privacy. Ralph et al. in [38] proposed a framework for de-identifying images that subsumes several previously introduced approaches. Their experiments used images from the CMU Multin PIE database and successfully de-identified using the nearest-neighbor classifier.

Newton et al. [49] presented a privacy-enabling k -Same algorithm that guarantees face recognition software cannot identify de-identified faces, even though many facial details are protected. Shai and Moshe presented an example of a face detection protocol that reveals no information to either party. The authors suggested an approach that avoids the need to de-identify images without the risk of revealing some information about the original image [32]. Willy et al. in [60] described a login system utilizing Two Factor Authentication and Zero-Knowledge Proof, which is used to maintain the confidentiality of the password, and Two Factor Authentication is used to secure the login process on untrusted devices. They tested their system, and initial results indicate that such a system can secure the login process without leaking the user's password. Cao et al. [42] implemented machine learning to detect fake accounts in this paper. They used a LinkedIn dataset, including the registration IP address and the registration date. Their model achieved an AUC of 0.95.

Michael et al. in [37] proposed a novel method to detect users who randomly connect to others. The algorithm has been evaluated on several social networks and has proven effective in detecting various types of malicious profiles. Ala et al. [17] In this work, they developed spam profile detection models based on a set of simple and publicly available Twitter features. They used four machine learning algorithms to develop the detection models and applied two feature selection methods. The best results in work obtained were using the Naive Bayes and Decision Trees classifiers. Table 1 presents a comparison between some researches that focus on security in the social network.

3 Social network

3.1 Types of social network

Online social networks are websites that allow users to connect with other internet users. It can be classified based on the nature of the services provided by these sites to [62]:

Table 1 Comparison between some researches that focus on security in the social network

Ref	Research title	Advantage	Limitation
Ilia et al. [41]	Face/Off: Preventing Privacy Leakage From Photos in Social Networks	They consider the face to be personally identifiable information (PII). When another user tries to access a photo, the system determines which faces the user does not have permission to see and blurs out the restricted faces. The system makes use of social networks' existing face recognition functionality and can communicate with existing photo-level access control mechanisms.	They effectively prevent users from identifying their contacts in 87.35%
Al Hasib [6]	Threats of Online Social Networks	This paper emphasizes the commercial and social benefits of safe and informed use of Social Network Sites, as well as the most significant threats to users of Social Network Sites, and the fundamental factors underlying these threats. It also makes policy and technical recommendations to improve privacy and security while maintaining the benefits of information sharing via Social Network Sites.	They have stated a few recommendations to enhance the security issues of Social Network Sites such as user awareness.
Gross and Sweeney [38]	Towards real-world face de-identification	They proposed a general framework for image de-identification that incorporates several previously introduced approaches.	they have focused on the identification of the image which is already used in several approaches and can be forged easily.
Jia et al. [42]	Random Walk Based Fake Account Detection in Online Social Networks	They implemented machine learning to detect fake accounts in this paper. They used a LinkedIn dataset, including the registration IP address and the registration date.	They have used one dataset only to detect fake accounts which is an Instagram dataset
Fire et al. [37]	Strangers Intrusion Detection - Detecting Spammers and Fake Profiles in Social Networks Based on Topology Anomalies	They proposed a novel method to detect users who randomly connect to others.	Their method was built based on random accounts which could include legitimate accounts
Our work	A Novel Machine Learning and Face Recognition Technique for Fake Accounts Detection System on Cyber Social Networks	we have used three different datasets with different features and nature, and we take advantage by using face recognition as a second-factor authentication.	

- Personal social networks: such as Facebook, where these networks enable users to create accounts and connect with other users. However, this network exchanges information and data with other individuals or applications.
- Status updates Social networks such as Twitter, where these networks allow users to post short status updates and communicate efficiently and quickly with others
- Location Social Network: These networks provide a way to reach one's position in real time, as a notification to authorized contacts or public knowledge. Example: Loopt.
- Shared interest Network: where such networks are built to facilitate the exchange of interests, like LinkedIn, where they share educational backgrounds and work experiences.
- Content Sharing Network: These networks are organized to share platforms, such as videos, songs, and pictures. Example: YouTube and Snapchat.

3.2 Architecture of OSNs

Online social networking sites serve individuals to communicate and build relationships with other social networking users. To achieve this goal, OSN has to include tools that facilitate users' interactions on the network, like registration, adding contacts, receiving notifications, and controlling the privacy of their accounts. The system of OSNs is formed by three layers: the data storage layer, the content management layer, and the application layer. Pallis et al. [53]

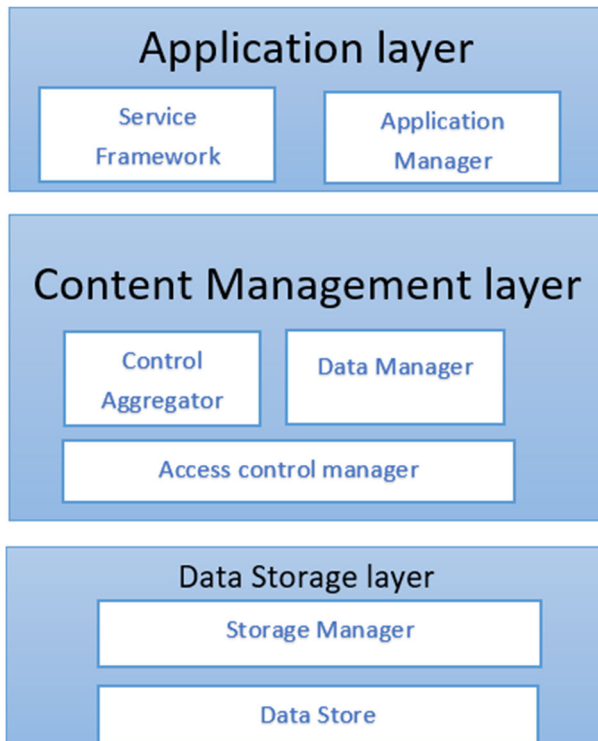


Fig. 2 Architecture of OSN

The data storage layer consists of 1. a storage manager, who is in charge of storing information and managing the rapid increase in database load through memory caching distribution. 2. A data storage component that stores social networking service information. Management Layer: This layer facilitates the conjunction of social information from remote OSN sites and the maintenance and recovery of the social content graph through the data manager. The application layer consists of a service framework and an application manager. Applications are usually delivered to the user by the application manager, facilitating interaction with users through a set of Application Programming Interfaces (APIs). Figure 2 shows the Architecture of the typical OSN.

3.3 Attacks on social network

OSN is an interaction-based application that allows registered users to interact with their friends and others in the network. These communications generate vast volumes of data which are shared through social networks. Attackers always have a head start on social engineering attacks compared to defenders. There are various types of threats [58], explained in this section as shown in Fig. 3.

3.3.1 Advanced persistent threats

Advanced persistent threats are a type of threat in which an attacker filters the identity of users and then uses their confidential information for malicious purposes [34]. The user who uses OSN usually submits his unique email address and provides confidential information like personal address, phone numbers, date of birth, information about his current location, and where he works. The attackers guess the registered user's password by analyzing the users' interests. Alternatively, maybe they will make a new account using the user's personal information that is available publicly. The attackers can employ different types of attacks, such as:

- A spear-phishing attack: where attackers gather the personal information of a user available on the OSN and use this information for the attack on other users by behaving as genuine users. The email sent by the attacker gains personal information about the user,

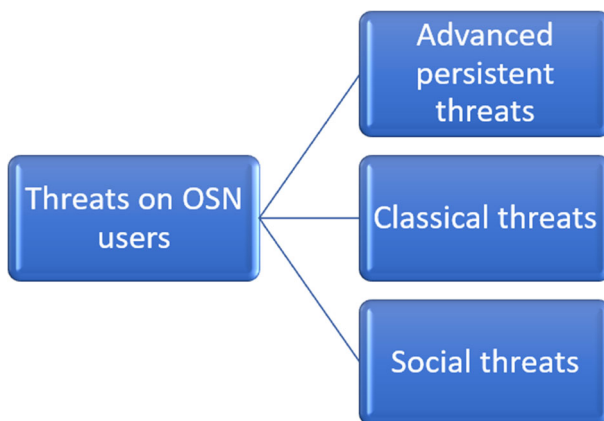


Fig. 3 Threats on OSN users

just like passwords and credit card information, and it is professional enough to seem like it is from the source [63].

- Fake profile attack: The attacker creates a new account on any social network that looks exactly like the original user's account and uses information already published about the genuine user to collect information from another user [59].
- DDoS attack: A DDoS attack is an attack coming from many locations, making a flood of requests, which overtakes users' resources and prevents users from serving legitimate customers [58].
- Online chat risk: While most OSNs allow users to communicate online through online chat. Users share their personal information during the chat, so the hacker can access their information and use these messages for malicious purposes [39].
- Illation attacks: This type of attack can be implemented by using network analysis, in which the attacker tries to guess the information about the user that is not available in the user's profile using available data [45].

3.3.2 Classical threats

This threat steals the original user's credentials and modifies the user's content. Classical threats can be classified into:

- Phishing attacks: Phishing is a type of social engineering attack often used to steal user data, which is used to access important accounts and can result in identity theft and financial loss. It occurs when an attacker, posing as a trusted, legitimate institution, dupes a victim through communication channels. The user is then lured into clicking a malicious link, which can cause the installation of malware, the freezing of the system as part of a ransomware attack, and the revealing of sensitive information.
- Malware: Malware is a malicious program. Different types of malware are classified; worms, viruses, trojans, backdoors, and ransomware. Recent research has shown the benefits of using social network analysis to classify the malware family. However, social networking functions such as degree distribution, degree centrality, average distance, clustering factor, and network density controlled by the graph structure of the malware system can be used to detect malware [56].
- Ransomware is a type of attack that blocks users until they pay money to the attacker, but there is no guarantee that the attacker will unblock them once they pay.
- Malicious URLs: The best way to disseminate exciting information in the OSN is through a website URL. The OSN user transmits information to his colleagues using website URLs. In an OSN, malicious URLs are spread through posts, comments, and likes. They can shorten URLs using various methods and features like bit.ly. Short URLs can hide original URLs and put suspicious websites behind short URLs. As a result, it can spread spam and phishing threats across the web [20].

3.3.3 Social threats

A social threat is a type of threat that an attacker uses to provoke people using social networking sites. They usually target young users or teenagers, making this threat severe and dangerous.

An attacker can gather their information through content-based access methods and then misuse it at a later stage to perform dangerous social engineering attacks. Corporate espionage can perform automated social engineering attacks with the help of OSN. Instead of

the classical social engineering approach, a social engineer uses OSN to collect personal information such as the employee's location and status, email ID, phone number, and more. [58]

3.4 Privacy issues in social networking websites

According to the seriousness of the threats mentioned above and the real effects of the attacks, the attention of social network operators and security companies has turned to deal with these threats through several solutions discussed in this section.

3.4.1 A. authentication mechanisms

OSNs use authentication mechanisms. In order to ensure that users who register or log in to social networks are real users and not attackers or stolen user accounts, OSN operators use authentication mechanisms, such as multi-factor authentication [60]. For example, Twitter recently launched its two-factor authentication mechanism [50], password and verification code sent to the user's mobile device. This mechanism can prevent malicious users from logging in using the hijacked account. This paper suggested a face-detected authentication method as a double factor authorization to enable authorized users to log in to their accounts using OSN, as no such method has been applied on social networks. The experiment is discussed in the next section in detail.

3.4.2 B. security and privacy settings

Privacy is a significant challenge in OSN, and several researchers have examined different aspects of the privacy issue. For example, studies show that Facebook's current privacy settings only match user expectations by 37% which indicates that current settings are often incorrect. It is a big concern because when the settings are incorrect, they almost always tend to be more open than the user's desired setting, exposing content to more users than expected. Although when users changed their default privacy settings, the modified settings only matched expectations 39% of the time, showing that even users who understood those who are more aware of privacy also have difficulty managing and maintaining their privacy settings correctly [44].

Some privacy methods that can be used on Facebook are that users can customize their privacy settings and choose which other users on the network can view their details and personal information. Some OSNs also support additional security configurations that allow the user to enable secure browsing, receive login notifications, and implement additional security features. The OSN, which still uses the default privacy settings, can be misused by others to use their information for malicious purposes.

3.4.3 C. internal protection mechanism

Some OSNs protect their users by establishing additional internal protection mechanisms to avoid such threats. For example, Facebook uses the immune system, which focuses on rapid detection and response, cross-channel data sharing, and integrated functional feedback loops. The Facebook immune system is a security infrastructure used by Facebook to detect spam and other online fraud. FIS uses smart software to detect suspicious links and behavioral patterns on social networking sites. A team of security experts oversees the software, but it can also learn and act on its own information [61].

3.4.4 D. report users

One of the solutions for security in social networks that protect users, especially teenagers, is using the option of reporting abuse or policy violations by other users [36].

4 Methodology

Our methodology is categorized into the following phases: Datasets Collection, Training Model Scoring, and Testing using machine learning classification techniques. Moreover, we proposed a classifier to classify faces based on the dataset collected from the live camera. The experimental phases are presented in detail in the following section.

4.1 Face processing digital authentication experiment

We need a dataset for training purposes for our deep learning project, so we thought using real data would give us the best accuracy for our model, which is to authenticate faces as a double factor authentication for social network security, as follows:

4.1.1 Collecting data

We used the webcam to capture one image each time for three faces and will take pictures where faces are present using the “vision.CascadeObjectDetector” Viola-Jones algorithm. We chose to train our model with 1207 total captures, categorized as the following: face1: “457”, face2: “300”, face3: “450” for each face. The webcam will break the loop when it reaches the maximum number captured. Then we copy and store the images for each face sub folder in our main dataset folder “database” with a “.bmp” extension, the detected images with a resizing of 227 x 227 as it is a requirement for AlexNet neural network transfer learning for image training, and if the faces were not detected, it would not be implemented. So we concluded with three face captures with a total of 1207 images in our database folder; this was our collection data part as shown in Table 2.

4.1.2 Training model

We used the AlexNet neural network deep learning algorithm in our training because it is a leading architecture for any object-detection task. It is a pre-trained model for 1000 images, but we will classify only three images in our model. We trained our model with

Table 2 Face processing dataset

Face No.	Captures
Face1	457
Face2	300
Face3	450
Total	1207

the stochastic gradient descent with momentum (SGDM) optimizer to classify our images based on our real dataset.

4.1.3 Scoring and testing

After scoring our trained model on our dataset, we achieved an accuracy of 95.31% which is accurate for deep learning classifying as we are not using a massive dataset as shown in the Fig. 4. Then we test our model based on our trained dataset to detect and classify the authorized face that has trained the model.

4.1.4 Suggested security approach

The data encryption method using the hash algorithm is suggested to achieve a higher level of security on the dataset in our “database” folder by adding a hash function and encrypting the whole dataset with the hash using the data owner’s public key. In this way, we can achieve identification, authentication, and integrity of the data to ensure that the collected data has not been altered and manipulated by the attacker over the internet.

4.2 Detection of fake profile in online social networks using machine learning

Fake accounts are the preferred method for attackers of online social networks to send spam, commit fraud, or just abuse the system. A single attacker may create dozens to thousands of fake accounts in order to raise their operation to reach the possible maximum number of legitimate members [42]. Detecting and taking action on these fake accounts as quickly as possible is essential to protect the actual legitimate members and preserve the trustworthiness of the network. However, the fake account may appear legitimate on first inspection by having a real-sounding name or a believable profile.

Our methodology is categorized into the following phases: Datasets Collection, Reprocessing, and Using machine learning classification techniques. We proposed a classifier to classify accounts based on the dataset features as legitimate and fake accounts.

Training on single CPU.

Initializing input data normalization.

```

=====
| Epoch | Iteration | Time Elapsed | Mini-batch |
|        |           | (hh:mm:ss)   | Accuracy   |
=====
| 1 | 1 | 00:00:14 | 23.44% |
| 3 | 50 | 00:12:14 | 85.94% |
| 6 | 100 | 00:24:36 | 95.31% |
    
```

Fig. 4 Face processing results

The first primary step is to have the required dataset; we have used two datasets that have been taken from available resources publicly. The reason for using two datasets with different features is to compare the result in different social networks and to identify how several features will affect the model's efficiency in detecting spam accounts. After collecting data, we reprocessed the datasets by removing the duplicated rows removing missing values to achieve the most accurate rate. After importing the processed dataset, we used two methods for training and validating the imported data splitting method into 70% train and 30% testing data. The other model uses the K-Folds method. This technique involves randomly dividing the dataset into k groups or approximately equal-sized folds. The first fold is kept for testing, and the model will be trained on k-1 folds. This process is repeated K times, and different folds are used for validation each time. The idea is that more training data makes the classification model more professional and accurate. The second phase is the training stage, as the classifier model, with the help of the selected Machine Learning algorithm, will be trained to classify the imported data into spam or legitimate accounts. The final stage is to validate the model's accuracy after completing the first and the second stages and classify the profiles into legitimate and spam accounts.

4.2.1 The used machine learning algorithms

To train and validate the accuracy of spam accounts detection [22, 24], eight supervised classification techniques were chosen based on their each different mechanisms of learning and handling with the dataset features as listed below:

1. Logistic Regression.
2. Naïve Bayes.
3. Support Vector Machine.
4. K-Nearest Neighbour (fine).
5. Boosted Tree (Ensemble).
6. Neural Network.
7. SVM kernel.
8. Logistic Regression Kernel.

4.2.2 Datasets description

The chosen datasets have been obtained from the public website [43], with the full wing details :

4.2.3 1. Facebook spam dataset

The first dataset, Facebook Spam Dataset, contains 500 legitimate and 100 spam profiles. The dataset features are as follows Label (0-legit, 1-spam).

- 1-Number of friends.
- 2-Number of followings.
- 3-Number of Community.
- 4-The age of the user account (in days).
- 5-Total number of posts shared.
- 6-Total number of URLs shared.
- 7-Total number of photos/videos shared.
- 8-Fraction of the posts containing URLs.

- 9-Fraction of the posts containing photos/videos.
- 10-Average number of comments per post.
- 11-Average number of likes per post.
- 12-Average number of tags in a post (Rate of tagging).
- 13-Average number of hashtags present in a post.

4.2.4 2.Instagram fake spammer genuine accounts

Instagram is one of the social media sites facing major problems with fakes and spam accounts, which can be detected using machine learning. This dataset comprises 696 instances, 348 legitimate accounts, and 348 fake ones. The 12 features in the chosen dataset were as follow:

1. profile pic.
2. nums/length username.
3. full name words.
4. nums/length full name.
5. name==username.
6. description length.
7. external URL.
8. private.
9. Numbers of posts.
10. Numbers of followers.
11. Numbers of follows.
12. label (fake =1, legitimate=0)

4.2.5 The experiments

4.2.6 Facebook's experiment

The first experiment was on the Facebook Spam dataset. This dataset consists of 14 features. We reprocessed it by deleting the duplicated rows and missing values before importing it to

Table 3 Accuracy results for facebook dataset

Algorithm	Accuracy with (5 Folds)	Accuracy With split dataset
Logistic Regression	95.7%	96.7%
Naïve Bayes	96.3%	97.2%
Support Vector Machine	96.3%	97.8%
K-Nearest Neighbour (Fine)	93.0%	94.4%
Boosted Tree (Ensemble)	87%	83.3%
Neural Network	93%	94.4%
SVM kernel	96.3 %	96.1%
Logistic Regression Kernel	94.5%	94.9%

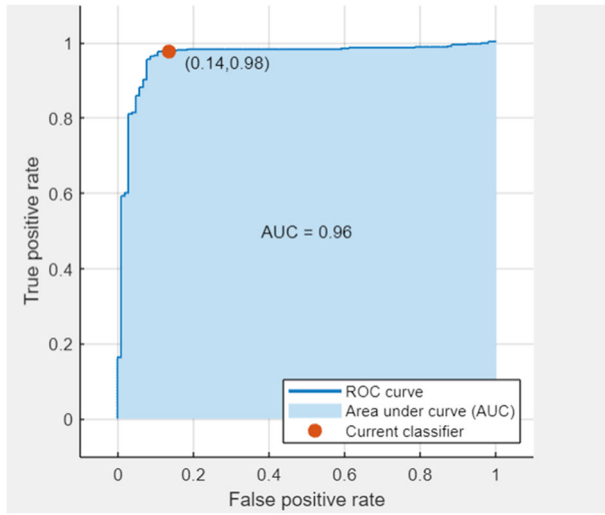


Fig. 5 Facebook logistic regression

our model to achieve the best accuracy rate. We applied our model to MATLAB, using the eight chosen ML algorithms, on a random sample of 500 legitimate and 100 spam accounts. We get our accuracy results by applying the model with two different techniques. Firstly, using K-fold (5). The second technique splits the dataset into 0.70 train and 0.30 test. The K-fold is used to evaluate machine learning models on a limited data sample. This process has a single parameter called k , usually equal to 1, 2, 5, or 10, and it refers to the number of groups a given data sample is to be split. The results for the Facebook dataset are represented in Table 3 :

As per the table above, it is clear that the results are not that different between the two methods. We can obtain that the SVM and Naive Bayes achieved the highest accuracy with

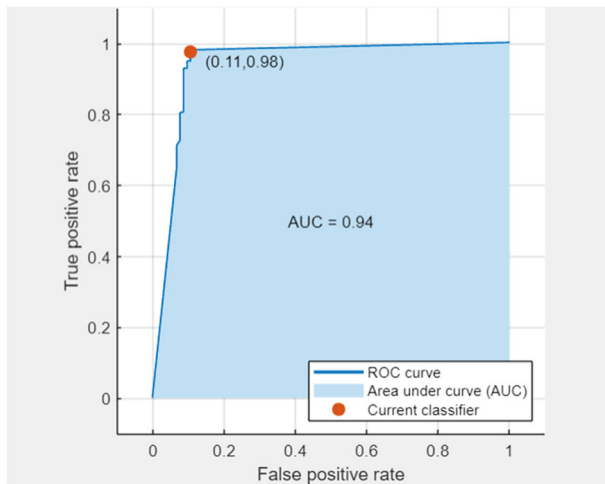


Fig. 6 Facebook-naïve bayes

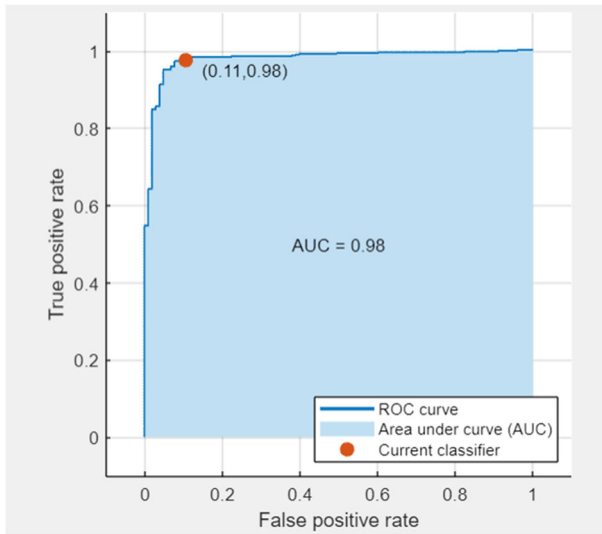


Fig. 7 Facebook-SVM

“96.3%” using the five folds CV method and with splitting with the average”97% and the lowest for the boosted tree with “87%” using five-fold CV and “83.3%” using splitting.

The Area Under the Curve (AUC) figures for our experiment; which indicate how successful a model is at separating positive and negative classes, are represented in Figs. 5, 6, 7, 8, 9, 10, 11 and 12 sequentially. The higher the AUC, the better the performance of the model.

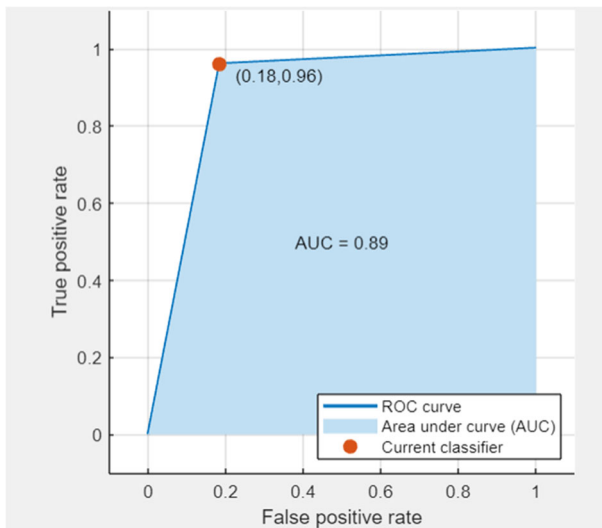


Fig. 8 Facebook-KNN

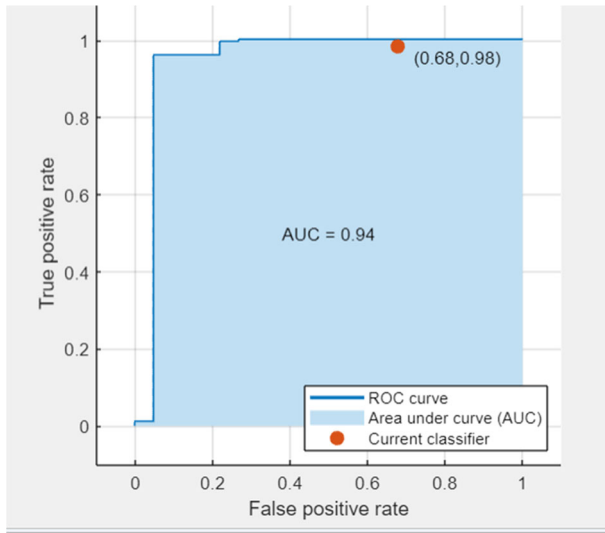


Fig. 9 Facebook-boosted tree

4.2.7 Instagram's experiment

The second experiment was on Instagram's fake spammer open accounts dataset. This dataset consists of 12 features. After applying our model through the Matlab App tool using the eight chosen ML algorithms mentioned previously, we get our accuracy results by applying the model in two different techniques. Firstly, using K-fold (5). The second technique splits the dataset into 0.70 train and 0.30 test. The results are represented in the Table 4:

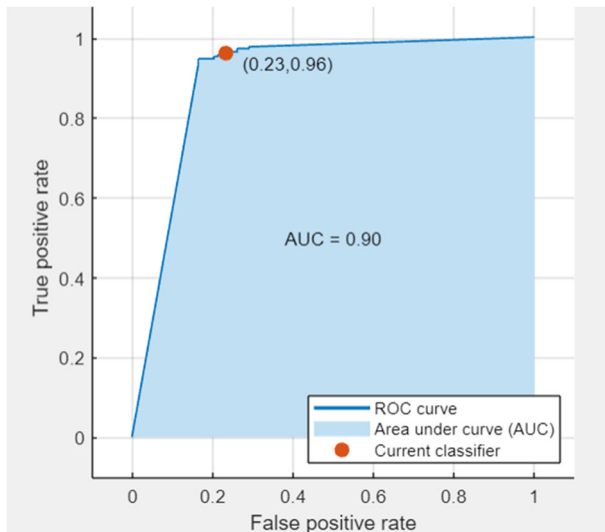


Fig. 10 Facebook-neural network

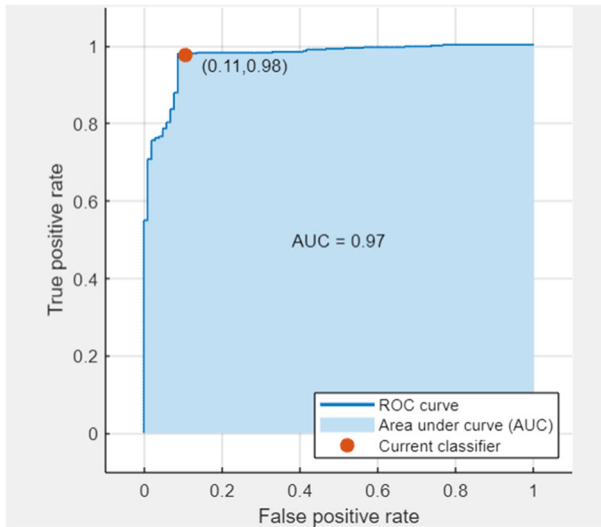


Fig. 11 Facebook-SVM kernel

As per the table above, it is clear that the results are not that different between the two methods. We can obtain that the Boosted Tree (Ensemble) achieved the highest accuracy with “94.7%”. The Area Under the Curve (AUC) figures for our experiment, which indicate how successful a model is at separating positive and negative classes, are represented sequentially in Figs. 13, 14, 15, 16, 17, 18, 19 and 20. The higher the AUC, the better the performance of the model.

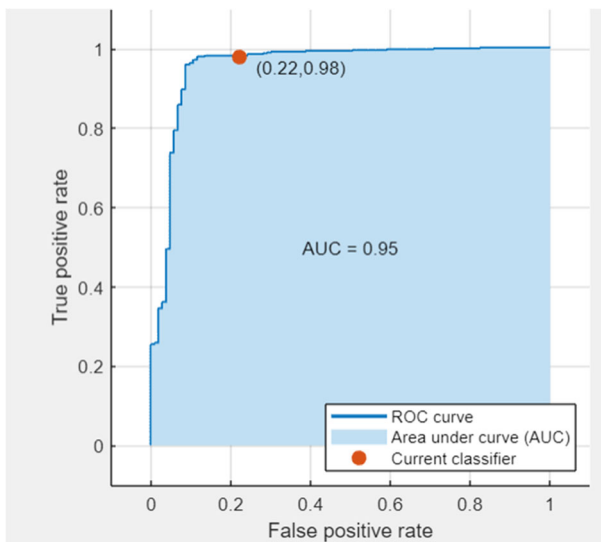
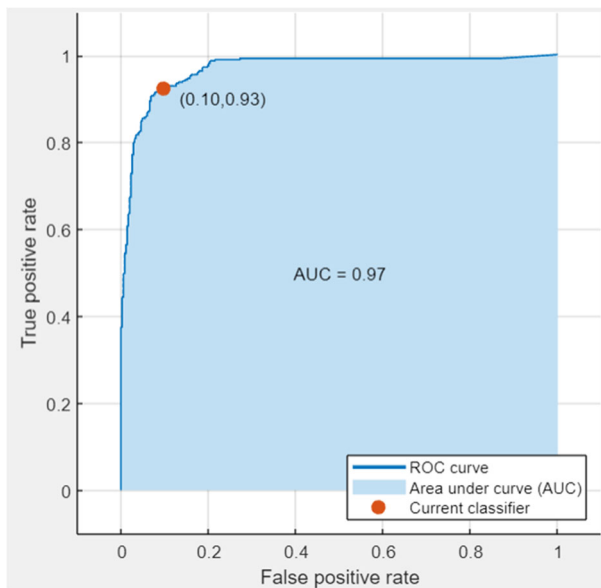


Fig. 12 Facebook-logistic regression kernel

Table 4 Accuracy results for instagram dataset

Algorithm	Accuracy with (5 Folds)	Accuracy With split dataset
Logistic Regression	91.4%	93.8%
Naïve Bayes	86.2 %	86.5%
SVM	89.4%	90.9%
KNN (fine)	87.4%	87.5%
Boosted Tree (Ensemble)	93.4%	94.7%
Neural Network	88.5%	88.9%
SVM kernel	80.6%	76.9%
Logistic Regression Kernel	80%	75.5%

Several researchers studied the impact of the number of instances on the performance of the detection model, and the results show the accuracy improved when the size of the sample increased. On the other hand, the time of detecting and computing increased as well with the increase of the sample size, which is considered as weak in the detecting model. In our experiment, although we have used a small dataset size, we have achieved a very good accuracy where SVM achieved the highest classification accuracy for the Fake Profiles detection datasets with 97.1 for future work, we will increase the sample size and study the effect of instances number on the performance. The classification in machine learning teaches us the security levels we need to maintain in social networking. In this research,

**Fig. 13** Instagram- logistic regression

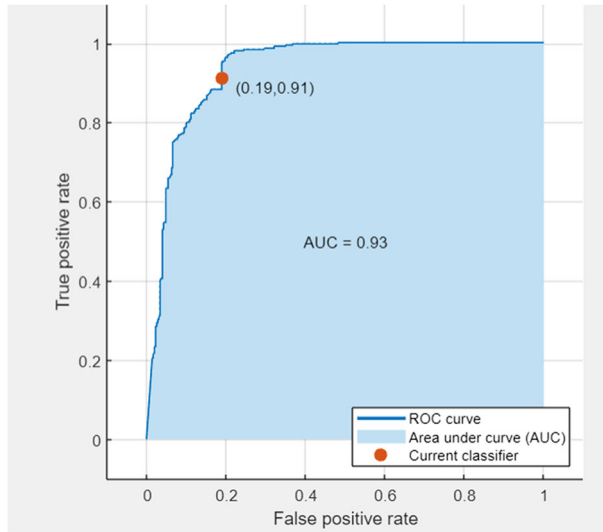


Fig. 14 Instagram Naive Bayes

we have used two different datasets; one for Facebook and the other for Instagram. The highest accuracy that we have is for SVM in the Facebook dataset with 97.8% and 93.8% using LR in the Instagram dataset. Interestingly, we noticed that the results are not that different between the two methods: simulating the dataset or using the K-folds we used. Table 5 presents a compression for our work with other researchers using ML to detect fake accounts on social networking accounts.

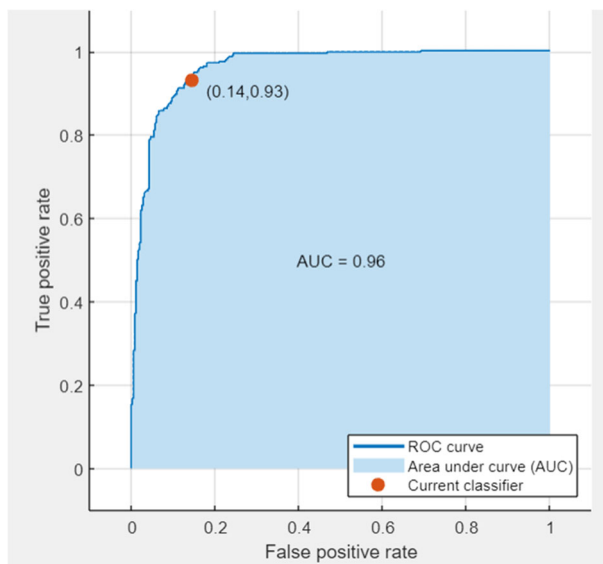


Fig. 15 Instagram-SVM

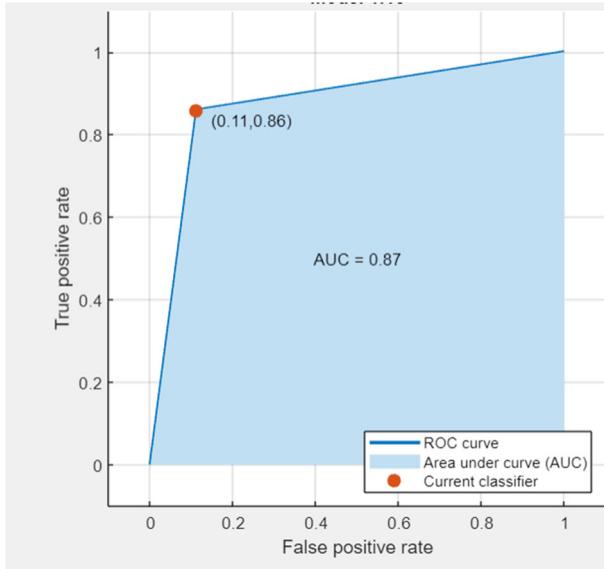


Fig. 16 Instagram- KNN

As we can see from our experiments and others, ML shows high accuracy in detecting fake accounts, although the experiments were applied on different social networks with different datasets and features. Although Instagram’s dataset was balanced with almost equal legitimate and fake accounts, it gives less accuracy than Facebook’s dataset. It almost returns to the features that were selected in each dataset.

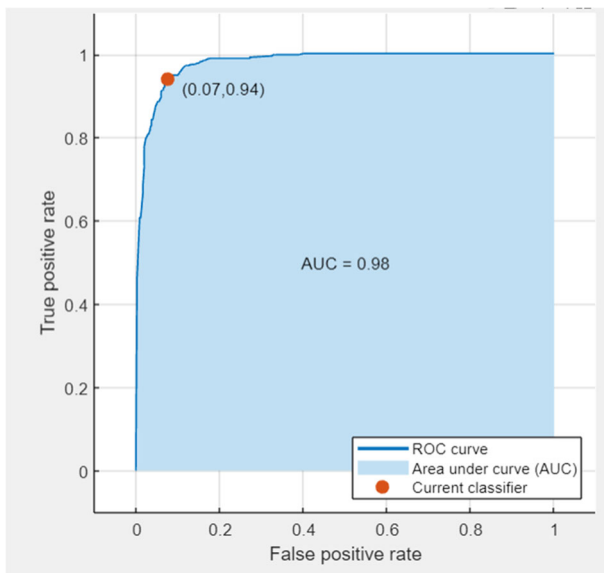


Fig. 17 Instagram- boosted tree

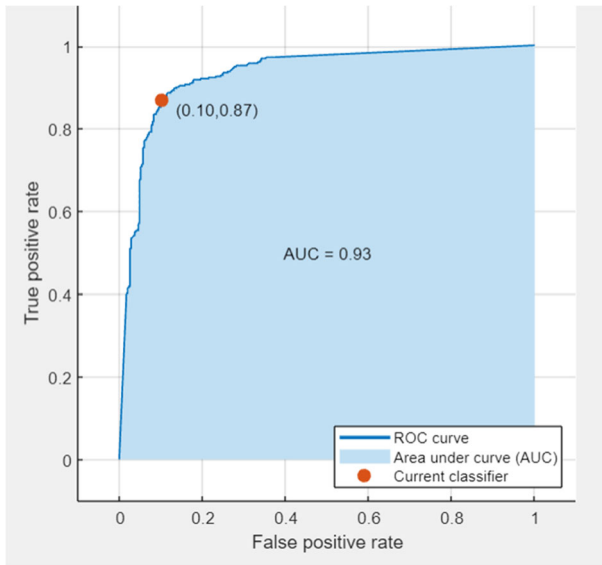


Fig. 18 Instagram- neural network

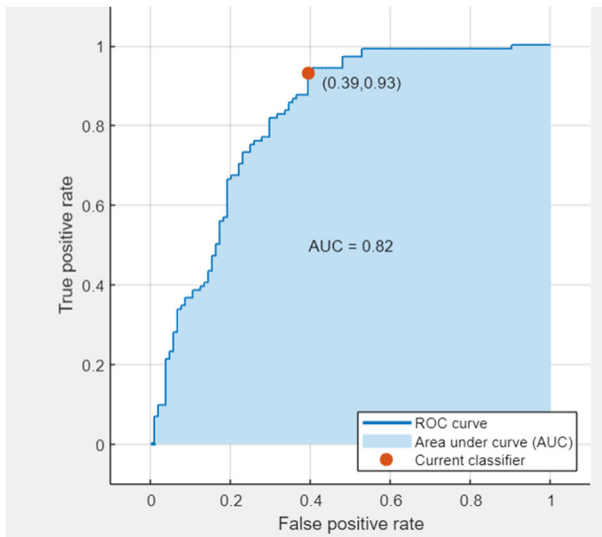


Fig. 19 Instagram- SVM kernel

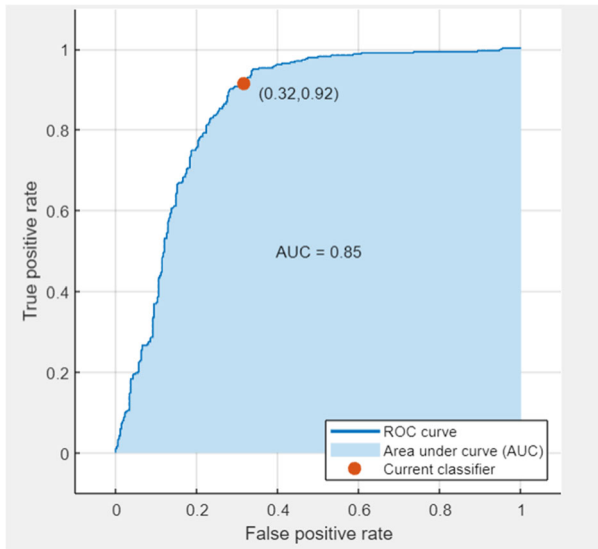


Fig. 20 Instagram- logistic regression kernel

Table 5 Researches used ML to detect fake account

Refrence	Socail networking type	ML algorithm	Accuracy
Raturi [55]	Facebook	SVM	97%
Raturi [55]	Inastegarm	SVM	97%
Raturi [55]	Twitter	SVM	99%
Raturi [55]	Youtube	SVM	99%
Raturi [55]	Whatsaap	SVM	89%
Aydin et al. [33]	Twitter	LR	79.1%
Albayati and Altamimi [18]	Facebook	SVM	95.27%
Albayati and Altamimi [18]	Facebook	KNN	91.40%
Albayati and Altamimi [19]	Facebook	SVM	95.72%

5 Conclusion and future work

In this article, we have presented the security and privacy issues on online social networks and a few of the suggested solutions based on researchers' studies. However, an ultimate solution is required for developing effective, secure, and privacy-preserving OSNs. Privacy is the most critical issue in OSNs since the unauthorized entry of users' private information can lead to disasters in all users' life sectors. The current method requires more effective security, especially with the increasing use of users. In order to be able to protect users' data, we applied a face detector model using a real dataset consisting of three faces. With a total of 1207 captures that have been taken from the live webcam, we used SGDM deep learning algorithms to classify the trained, authorized faces. The accuracy rate we achieved was 95.31%. Since no such method has been applied as a second-factor authentication, we propose a detection model for fake spam accounts in OSN based on two chosen datasets which consist of legitimate and fake accounts from Facebook and Instagram social networks. Using MATLAB, we applied ML algorithms after applying two techniques for the training phase: splitting the data into 0.70 for training and the remaining 0.30 for validating, and using the k-fold technique; the result we obtained shows that machine learning can investigate high accuracy in detecting fake accounts. For future work, we suggest increasing the face recognition dataset to compare the results with our current sample, which achieved high accuracy results.

Data Availability Data is available from the authors upon reasonable request.

Declarations

Conflict of Interests The authors declare that there is no conflict of interest regarding the publication of this paper.

Ethics approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Abualigah L, Alfar HE, Shehab M, Hussein AMA (2020) Sentiment analysis in healthcare: a brief review. *Recent Adv NLP Case Arabic Lang*:129–141
2. Abualigah L, Diabat A (2022) Chaotic binary reptile search algorithm and its feature selection applications. *J Ambient Intell Humanized Comput*:1–17
3. Abualigah L, Kareem NK, Omari M, Elaziz MA, Gandomi AH (2021) Survey on twitter sentiment analysis: architecture, classifications, and challenges. In: *Deep learning approaches for spoken and natural language processing*. Springer, pp 1–18
4. Abushanap SA, Abdalla AM, Tamimi AA, Alzu'bi S (2021) A survey of human face recognition for partial face view. In: *2021 International conference on information technology (ICIT)*, IEEE, pp 571–576
5. Abusukhon A, AlZu'bi S (2020) New direction of cryptography: a review on text-to-image encryption algorithms based on rgb color value. In: *2020 Seventh international conference on software defined systems (SDS)*, IEEE, pp 235–239
6. Al Hasib A (2009) Threats of online social networks. *IJCSNS Int J Comput Sci Netw Security* 9(11):288–93
7. Al-Arjan A, Rasmi M, AlZu'bi S et al (2021) Intelligent security in the era of ai: the key vulnerability of rc4 algorithm. In: *2021 International conference on information technology (ICIT)*, IEEE, pp 691–694

8. Al-Zu'bi S, Hawashin B, Mughaid A, Baker T (2021) Efficient 3d medical image segmentation algorithm over a secured multimedia network. *Multimed Tools Appl* 80(11):16887–16905
9. AlZu'bi S, Abu Zitar R, Hawashin B, Abu Shanab S, Zraiqat A, Mughaid A, Almotairi KH, Abualigah L (2022) A novel deep learning technique for detecting emotional impact in online education. *Electronics* 11(18):2964
10. AlZu'bi S, Alsmadiv A, AlQatawneh S, Al-Ayyoub M, Hawashin B, Jararweh Y (2019) A brief analysis of amazon online reviews. In: 2019 Sixth international conference on social networks analysis, management and security (SNAMS). IEEE, pp 555–560
11. AlZu'bi S, Alsmirat M, Al-Ayyoub M, Jararweh Y (2019) Artificial intelligence enabling water desalination sustainability optimization. In: 2019 7th International renewable and sustainable energy conference (IRSEC), IEEE, pp 1–4
12. AlZu'bi S, Aqel D, Lafi M (2022) An intelligent system for blood donation process optimization-smart techniques for minimizing blood wastages. *Clust Comput*:1–11
13. AlZu'bi S, Aqel D, Mughaid A, Jararweh Y (2019) A multi-levels geo-location based crawling method for social media platforms. In: 2019 Sixth international conference on social networks analysis, management and security (SNAMS), IEEE, pp 494–498
14. AlZu'bi S, Hawashin B, ElBes M, Al-Ayyoub M (2018) A novel recommender system based on apriori algorithm for requirements engineering. In: 2018 Fifth international conference on social networks analysis, management and security (snams). IEEE, pp 323–327
15. AlZu'bi S, Hawashin B, Mujahed M, Jararweh Y, Gupta BB (2019) An efficient employment of internet of multimedia things in smart and future agriculture. *Multimed Tools Appl* 78(20):29581–29605
16. AlZu'bi S, Shehab M, Al-Ayyoub M, Jararweh Y, Gupta B (2020) Parallel implementation for 3d medical volume fuzzy segmentation. *Pattern Recogn Lett* 130:312–318
17. Ala'M A-Z, Alqatawna J, Paris H (2017) Spam profile detection in social networks based on public features. In: 2017 8th International conference on information and communication systems (ICICS). IEEE, pp 130–135
18. Albayati M, Altamimi A (2019) Mdfp: a machine learning model for detecting fake facebook profiles using supervised and unsupervised mining techniques. *Int J Simulation Syst Sci Technol* 20(1):1–10
19. Albayati MB, Altamimi AM (2019) Identifying fake facebook profiles using data mining techniques. *J ICT Res Appl*, vol 13(2)
20. Alghamdi B, Watson J, Xu Y (2016) Toward detecting malicious links in online social networks through user behavior. In: 2016 IEEE/WIC/ACM international conference on web intelligence workshops (WIW). IEEE, pp 5–8
21. Ali A, Zhu Y, Chen Q, Yu J, Cai H (2019) Leveraging spatio-temporal patterns for predicting citywide traffic crowd flows using deep hybrid neural networks. In: 2019 IEEE 25th international conference on parallel and distributed systems (ICPADS), IEEE, pp 125–132
22. Ali A, Zhu Y, Zakarya M (2021) Exploiting dynamic spatio-temporal correlations for citywide traffic flow prediction using attention based neural networks. *Inf Sci* 577:852–870
23. Ali A, Zhu Y, Zakarya M (2021) A data aggregation based approach to exploit dynamic spatio-temporal correlations for citywide crowd flows prediction in fog computing. *Multimed Tools Appl* 80(20):31401–31433
24. Ali A, Zhu Y, Zakarya M (2022) Exploiting dynamic spatio-temporal graph convolutional neural networks for citywide traffic flows prediction. *Neural networks* 145:233–247
25. Alia MA, Hnaif AA, Al-Anie HK, Maria KA, Manasrah AM, Sarwar MI (2011) A novel header matching algorithm for intrusion detection systems, arXiv:1108.1417
26. Alia MA, Maria KA, Alsarayreh MA, Maria EA, Almanasra S (2019) An improved video steganography: using random key-dependent. In: 2019 IEEE Jordan international joint conference on electrical engineering and information technology (JEEIT), IEEE, pp 234–237
27. Alkhatib AA, Abu Maria K, Alzu'bi S, Abu Maria E (2022) Novel system for road traffic optimisation in large cities. *IET Smart Cities*
28. Alkhatib AA, Alsabbagh A, Maraqa R, Alzubi S (2021) Load balancing techniques in cloud computing: extensive review. *Adv Sci Technol Eng Syst J* 6(2):860–870
29. Alzu'bi S, Jararweh Y (2020) Data fusion in autonomous vehicles research, literature tracing from imaginary idea to smart surrounding community. In: 2020 Fifth international conference on fog and mobile edge computing (FMEC), IEEE, pp 306–311
30. Alzubi S, Hawashin B, Mughaid A, Jararweh Y (2020) Whats trending? an efficient trending research topics extractor and recommender. In: 2020 11th International conference on information and communication systems (ICICS), IEEE, pp 191–196
31. Aqel D, Al-Zubi S, Mughaid A, Jararweh Y (2022) Extreme learning machine for plant diseases classification: a sustainable approach for smart agriculture. *Clust Comput* 25(3):2007–2020


32. Avidan S, Butman M (2006) Blind vision. In: European conference on computer vision. Springer, pp 1–13
33. Aydin I, Mehmet S, Salur MU (2018) Detection of fake twitter accounts with machine learning algorithms. In: 2018 International conference on artificial intelligence and data processing (IDAP), IEEE, pp 1–4
34. Bilge L, Strufe T, Balzarotti D, Kirda E (2009) All your contacts are belong to us: automated identity theft attacks on social networks. In: Proceedings of the 18th international conference on World wide web, pp 551–560
35. Elbes M, Kanan T, Alia M, Ziad M (2022) Covid-19 detection platform from x-ray images using deep learning. *Int J Adv Soft Comput Appl*, vol 14(1)
36. Fire M, Goldschmidt R, Elovici Y (2014) Online social networks: threats and solutions. *IEEE Commun Surveys Tutorials* 16(4):2019–2036
37. Fire M, Katz G, Elovici Y (2012) Strangers intrusion detection-detecting spammers and fake profiles in social networks based on topology anomalies. *Human J* 1(1):26–39
38. Gross R, Sweeney L (2007) Towards real-world face de-identification. In: 2007 First IEEE international conference on biometrics: theory, applications, and systems. IEEE, pp 1–8
39. Humphreys L (2007) Mobile social networks and social practice: a case study of dodgeball. *J Comput-Mediated Commun* 13(1):341–360
40. Hussein F, Mughaid A, AlZu'bi S, El-Salhi SM, Abuhajja B, Abualigah L, Gandomi AH (2022) Hybrid clahe-cnn deep neural networks for classifying lung diseases from x-ray acquisitions. *Electronics* 11(19):3075
41. Ilija P, Polakis I, Athanasopoulos E, Maggi F, Ioannidis S (2015) Face/off: preventing privacy leakage from photos in social networks. In: Proceedings of the 22nd ACM SIGSAC conference on computer and communications security, pp 781–792
42. Jia J, Wang B, Gong NZ (2017) Random walk based fake account detection in online social networks. In: 2017 47th Annual IEEE/IFIP international conference on dependable systems and networks (DSN). IEEE, pp 273–284
43. Aydin I, Mehmet SEVİ, Salur MU (2022) Detection of fake twitter accounts with machine learning algorithms. In: 2018 International Conference on Artificial Intelligence and Data Processing (IDAP) (pp. 1–4). IEEE
44. Liu Y, Gummadi KP, Krishnamurthy B, Mislove A (2011) Analyzing facebook privacy settings: user expectations vs. reality. In: Proceedings of the 2011 ACM SIGCOMM conference on internet measurement conference, pp 61–70
45. Mislove A, Viswanath B, Gummadi KP, Druschel P (2010) You are who you know: inferring user profiles in online social networks. In: Proceedings of the third ACM international conference on Web search and data mining, pp 251–260
46. Mughaid A, Al-Zu'bi S, Al Arjan A, Al-Amrat R, Alajmi R, Zitar RA, Abualigah L (2022) An intelligent cybersecurity system for detecting fake news in social media websites. *Soft Comput* 26(12):5577–5591
47. Mughaid A, AlZu'bi S, Alnajjar A, AbuElsoud E, Salhi SE, Igried B, Abualigah L (2022) Improved dropping attacks detecting system in 5g networks using machine learning and deep learning approaches. *Multimed Tools Appl*:1–23
48. Muhairat M, ALZu'bi S, Hawashin B, Elbes M, Al-Ayyoub M (2020) An intelligent recommender system based on association rule analysis for requirement engineering. *J Univ Comput Sci* 26(1):33–49
49. Newton EM, Sweeney L, Malin B (2005) Preserving privacy by de-identifying face images. *IEEE Trans Knowl Data Eng* 17(2):232–243
50. O'Leary J (2013) Getting started with login verification. *Twitter Blogs*
51. Obeidat I, Mughaid A, Alzoubi S (2016) A secure encrypted protocol for clients' handshaking in the same network. *Int J Interactive Mobile Technol (ijIM)*
52. Otair M, Ibrahim OT, Abualigah L, Altalhi M, Sumari P (2022) An enhanced grey wolf optimizer based particle swarm optimizer for intrusion detection system in wireless sensor networks. *Wirel Netw* 28(2):721–744
53. Pallis G, Zeinalipour-Yazti D, Dikaiakos MD (2011) Online social networks: status and trends. *New Direct Web Data Manag* 1:213–234
54. Ramadan R, Alqatawneh S, Ahalaiqa F, Abdel-Qader I, Aldahoud A, AlZoubi S (2019) The utilization of whatsapp to determine the obsessive-compulsive disorder (ocd): a preliminary study. In: 2019 Sixth international conference on social networks analysis, management and security (SNAMS), IEEE, pp 561–564
55. Raturi R (2018) Machine learning implementation for identifying fake accounts in social network. *Int J Pure Appl Math* 118(20):4785–4797

56. Reddy V, Kolli N, Balakrishnan N (2021) Malware detection and classification using community detection and social network analysis. *J Comput Virology Hacking Tech*:1–14
57. Safaldin M, Otair M, Abualigah L (2021) Improved binary gray wolf optimizer and svm for intrusion detection system in wireless sensor networks. *J Ambient Intelligence Humanized Comput* 12(2):1559–1576
58. Sahoo SR, Gupta BB (2019) Classification of various attacks and their defence mechanism in online social networks: a survey. *Enterprise Inf Syst* 13(6):832–864
59. Siddiqui S, Khan MS, Ferens K, Kinsner W (2016) Detecting advanced persistent threats using fractal dimension based machine learning classification. In: *Proceedings of the 2016 ACM on international workshop on security and privacy analytics*, pp 64–69
60. Song A (2011) Introducing login approvals. Tersedia: <https://www.facebook.com/notes/facebook-engineering/introducinglogin-approvals/10150172618258920>. Accessed 1 Jan 2021
61. Stein T, Chen E, Mangla K (2011) Facebook immune system. In: *Proceedings of the 4th workshop on social network systems*, pp 1–8
62. Sushama C, Kumar MS, Neelima P (2021) Privacy and security issues in the future: a social media. *Materials today: proceedings*
63. Williams EJ, Hinds J, Joinson AN (2018) Exploring susceptibility to phishing in the workplace. *Int J Human-Comput Studies* 120:1–13
64. Zhang Z, Gupta BB (2018) Social media security and trustworthiness: overview and new direction. *Futur Gener Comput Syst* 86:914–925

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Affiliations

Ala Mughaid¹ · Ibrahim Obeidat¹ · Shadi AlZu'bi² · Esraa Abu Elsoud¹ · Asma Alnajjar¹ · Anas Ratib Alsoud³ · Laith Abualigah^{4,5,3,6,7,8} 

¹ Department of Information Technology, Faculty of Prince Al-Hussien bin Abdullah for IT The Hashemite University, P.O. Box 330127, Zarqa, Jordan

² Faculty of Science and IT, Al-Zaytoonah University of Jordan, Amman, Jordan

³ Hourani Center for Applied Scientific Research, Al-Ahliyya Amman University, Amman, 19328, Jordan

⁴ Center for Engineering Application & Technology Solutions, Ho Chi Minh City Open University, Ho Chi Minh, Viet Nam

⁵ Computer Science Department, Prince Hussein Bin Abdullah Faculty for Information Technology, Al al-Bayt University, Mafrq 25113, Jordan

⁶ Faculty of Information Technology, Middle East University, Amman 11831, Jordan

⁷ Applied Science Research Center, Applied Science Private University, Amman 11931, Jordan

⁸ School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang 11800, Malaysia