Preventing Fake Accounts on Social Media Using Face Recognition Based on Convolutional Neural Network

Vernika Singh, Raju Shanmugam, and Saatvik Awasthi

Abstract In today's world, most people are intensely dependent on online social networks (OSN). People use social sites to find and make friends, to associate with people who share comparable intrigue, trade news, organize the event, exploring passion. According to a Facebook review, 5% of monthly active users had fake accounts, and in the last six months, Facebook has deleted 3 billion accounts. According to the Washington Post, Twitter has suspended over 1 billion suspect accounts over a day in recent months. Detection of a fake profile is one of the critical issues these days as people hold fake accounts to slander image, spread fake news, promote sarcasm that has attracted cybercriminals in. There are numerous machine learning methodologies, such as supervised learning and SVM-NN, and they are produced for the effective detection of a fake profile. In this paper, convolutional neural networks are proposed with many artificial neural network algorithms like face recognition, prediction, classification, and clustering for the efficient identification of account being real or fake and elimination of fake profile account. Furthermore, the study is grounded on the fact of the face recognizing of the user and performing feature detection and time series prediction. If the user account detected fake, it would not be created.

Keywords Social media analytics · Online social media · Face recognition · Convolutional neural networks · Age prediction

V. Singh (✉) · S. Awasthi
Galgotias University, Greater Noida, India
e-mail: vernikasingh8@gmail.com

S. Awasthi
e-mail: saatvikawasthi1998@gmail.com

R. Shanmugam
School of Computing Science & Engineering, Galgotias University, Greater Noida, India
e-mail: dr.sraju@yahoo.com

P. Karuppusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 55,
https://doi.org/10.1007/978-981-15-8677-4_19

1 Introduction

There are increasing numbers of people in the current generation with social online networks like Facebook, Twitter, Instagram, LinkedIn, and Google + [1]. Social networks allow people with common interests or collaborative reasons. It provides them with access to numerous services for example messaging, posting comments on their cyberwalls which are public, commenting on other users' profiles post, and exchanging the masking of identity for malicious purposes has become progressively prevalent over the last few years. People rely heavily on OSNs to remain in contact, to organize trade news, activities, and perhaps even e-business [2]. People rely heavily on online social networks (OSNs) to create and share individual personal profiles, email, images, recordings, audios, and videos, and to find and make friends that have attracted cybercriminals' interest in carrying out a variety of malignant activities. Government associations are used (OSNs) as a forum for effectively providing government-driven services to people and educating and informing them about different situations. This heavy utilization of social networks results in immense measures of data being disseminated and organizations use social networks to promote, advertise, and support their business online. Fake profile accounts show that people do not represent them as a real person. Such an account is manually opened by a person after that actions are automated by bot. Fake profile account is categorized into a Sybil account and duplicate account. A duplicate account applies to a user's account and maintained by the user other than their main account. Fake accounts are classed into user classified (reported) or unauthorized (unwanted) groups of accounts. User malicious account records show individual profiles made by a client from a company or non-human element. Alternatively, undesirable accounts are however user identities that are configured to be used for violation of security and privacy.

The social networking site database for Facebook records a statistic of 4.8% for duplicate accounts, the number of user-misclassified accounts is 2.4%, and the number of unauthorized accounts is 1.5% [3]. In 2019, Facebook announced the deletion of 2.3 billion fake profile accounts. This is almost twice as many as 1.2 billion accounts withdrawn in the first quarter of 2018. The Facebook Compliance Report shows that as much as 5 million of its monthly active users are fake and that there is a growing number of attacks. Facebook is estimated to have more than two billion monthly active users and one billion active users each day in the major online social network reports. Accordingly, only 5% of its active monthly users are false in Facebook reports [4]. It is very convenient today to make false accounts. Nowadays, fake profile accounts can be purchased on the Web at an extremely cheaper cost; furthermore, it can be delivered to the client using publicly supporting. Now it is easier to purchase followers online from Twitter and Instagram. The goal behind Sybil account formation is to defame someone else's image, digital terrorism, terrorist propaganda, fear-based oppressor publicity, campaigns for radicalization, distribution of pornography, fraud and misinformation, popularity shaping, division of opinions, identity insecurity. In this paper, the promptly accessible and designed methodologies are evaluated that are utilized for the fruitful detection of identifying fake accounts on Facebook utilizing AI models, of human-created detection that is accomplished by observing the attitude and the needs of people by examining their experiences. Detection is accomplished by observing the attitude and the needs of people by examining their communication and interaction with each other. Then, there is a need to give a fake account to our machine learning model to allow the algorithm to comprehend what a fake account is. Convolutional neural network (CNN) and feature classification algorithms are being used. Our methodology is to differentiate on the reality between true user accounts and fake accounts by checking them at the time of creation and not allowing the fake profiles to be created.

2 Problem Identification

2.1 Online Social Network (OSN)

Social media play a vital role in our life as 45% of the population around the world spend at least one hour daily on social networks to share news, post pictures, tweeting, commenting, liking, etc. [1]. Companies use the online social network to advertise and promote their business online. Government organization uses social media as a platform to deliver government services to citizens in an efficient manner and to educate and inform them in various ways. The highly dependent nature of billions of people over social networks has attracted the interest of cybercriminals to carry out malicious activities.

Computer-mediated communication (CMC) is a basic portion of our every day lives as, a result it has ended up essential for the populace, it is never going stalwart, as the society is never returning to a more physical environment, to considering characters on CMC stages like social media and gatherings. The idea of identity is expressly distinguished in this consider because it has to be overhauled logically as a critical theme of the social media wrangle about. As expressed prior, a few analysts have examined this issue, but determined three-discusses are required since it is connected to "self" and "self" as demonstrated overnight. Inquiry about concerned with the identity, picture, and online gatherings has numerous possibilities and is especially meriting of thorough consideration owing to social organizing designs predominant over thousands of a long time. Amid the final era, numerous more youthful people effectively lead complicated genuine organizing, which may not reflect their current or offline presence. It has ended up a matter of concern. As of late specified viewpoint is a fair one in setting up untrue characters in social media. In, however, another neighborhood, faux accounts are made in advanced organizing and reasonable consideration was done that has moreover procured.

2.2 Fake Profile

Fake profiles are created to defame someone's image by using an individual name and pictures, and terrorist propaganda, sharing fake news, and distribution of pornography. It has a high impact on younger age who are highly dependent on social networks. Most people try to portray an image of themselves who they are not and try to communicate with their dear ones and hate ones to harm them mentally which lead to serious problems such as depression or suicide. Fake accounts can be categorized into duplicate account which is maintained by a user in addition to their principal account whereas false account further breaks down into user-misclassified accounts which represent the user personal account created as a business account and undesirable accounts which are created to carry out malicious activities.

Social networking site, Facebook, has estimated that 4.8% profiles are double accounts, 2.3% profiles are unsolicited, and 1.6% are undesirable accounts [3]. The Facebook company has started removing 2.2 billion accounts that were detected fake in the Q1of 2019. That is double the number detected in Q4 of 2018 where 1.2 billion fake profiles were deleted. According to Facebook's Enforcement Report, it is 6% monthly active users are fake and the increase is due to the rise in automated attacks.[5] The creation of fake accounts and buying Twitter and Instagram followers along with putting likes online on posts are very easy nowadays. It can be bought online at a very less cost and can be given to the customer via crowdsourcing services.

The reason behind the creation of fake accounts is mostly among these, i.e., either to defame another person, terrorist propaganda, spreading rumors and false news, influencing popularity, polarizing opinions, identity theft, radicalization campaigns, cyberbully, dissemination of pornography and fraud, and online extremism.

2.3 Approach to Solve

The paper finds out the methods and the measures currently available for the detection of counterfeit accounts using machine learning technologies. Detection is carried out by learning the behaviors of individuals and requires a detailed analysis of the activities of social media interactions with other accounts. To identify other fake accounts and to predict fake accounts for potential use, this machine learning algorithm is training with the latest collection of identified fake profiles. Neural networks and algorithms are used for classification. Our solution uses distinct characteristics of the platform and restricts a platform to use an identity over the so-called Internet.

3 Literature Survey

This section summarizes some of the related work done in the field of detecting fake profile accounts using machine learning models.

3.1 Sarah Khaled, Neamat El-Tazi and Hoda M. O. Mokhtar

In 2018, a new algorithm is proposed to detect fake accounts and bots on Twitter. Classification algorithms of machine learning have been utilized, and they focus on technologies adopted to detect fake accounts and bots on Twitter. Classification algorithms of machine learning have been utilized to choose real or fake target accounts, and those techniques were support vector machine (SVM) and neural network (NN). They proposed new algorithms support vector machine and neural networks for successful detection of fake accounts. Both approaches adopt techniques of machine learning which highlight collection and methodologies of data reduction. It was also noted that the correlation collection records quality is dynamic among the other optimization algorithms, as redundancy is eliminated. The new algorithm classified 98% of the account of training dataset using fewer features [4].

3.2 Mudasir Ahmad wania, Nancy Agarwala & Suraiya Jabinb Syed Zeeshan Hussainb

In 2018, main focus is on fake profile detection using sentiments. The study is done on the post of real account user and fake profile user and similar emotions they use. The experiment is done on Facebook user profile post. In this paper, the author mainly focuses on fake profile detection using sentiments. The study is done on the post of real account user and fake profile user and similar emotions they use. The experiment is done on Facebook user profile post. Data are trained for 12 emotions, including 8 basic emotions, positive and negativeness, by the use of machine training techniques consequently, outliers are removed using noise removal technique. To train the detection model, many learning algorithms have been used including support vector machine (SVM), multilayer perceptron, J Rip, and random forest. The result shows that in the posts of unverified accounts three types of feeling, anxiety, shock, and faith are found least. For all three measures, precision, estimation, and AUROC, random forest provides the best result [5].

3.3 Estée van der Walt and Jan Eloff

Described the detection of fake identities of humans vs bots using machine learning models. Numerous fake accounts are enhanced with features used to detect bot accounts, and the collection of features has been extended to different supervised learning models. This paper focuses on the detection of fake identities of humans vs bots using machine learning models. Numerous fake accounts are enhanced with features used to detect bot accounts, and the collection of features has been extended to different supervised learning models. The highlights of human and machine accounts are indistinguishable. For occasion: The title it illustrates that the traits utilized to recognize programmer accounts fizzled to recognize human account points of interest appropriately. The effects of qualified computer models are predictive 49.75% of the best F1 performance. This is due to the fact that human beings are distinct from both in terms of behavior and characteristic, which cannot be modeled in the same way [6].

3.4 Gayathri A, Radhika S & Mrs. Jayalakshmi S. L.

In 2018, identification of fake accounts in media application by using support vector machines and neural networks is explained. Problem definition: identification of fake accounts in media application by using support vector machines and neural networks. In this report, they reflect a profound learning pipeline for identifying fake accounts rather than utilizing presumptions. It classifies the Sybil account cluster whether it is made by the same individual. The method starts by selecting a profile, at that point extricating the fitting characteristics and passing them to a proficient classificatory that categorizes the account as untrue or veritable alongside the input. Future work: utilizing more complex calculations. The other work line is to mimic multimodels utilizing the chosen highlights of other malware-based methods [7].

Author—Year	Objective	Techniques used	Accuracy
Naman Singh, Tusshar Sharma,Abha Thakral, Tanupriya Choudhury—2018	Detection of fake accounts	Support vector machine–neural networks	93%
Gayathri A, Radhika S, Mrs. Jayalakshmi	Detecting fake accounts in media application	Support vector machine & deep neural networks	–
Mudasir Ahmad wania, Nancv Agarwala, Suiaiya Jabinb, SyedZeeshan Hussain—2019	Analysis of real and fake users in Facebook based on emotions	Naive Bayes, J Rip, random forest	Random forest

(continued)

(continued)

Author—Year	Objective	Techniques used	Accuracy
Mehmet Şimşek, Oğuzhan Yılmaz, AsenaHazal Kahriman, Levent Sabah—2015	Detecting fake Twitter accounts	Artificial neural networks	–
Oscar S. Siordia, Daniela Moctezuma—2016	Features combination for the detection of malicious Twitter accounts	Feature extraction classification. Random forest	94%
Dr. Vijay Tiwar—2017	Analysis and detection of fake profile	Honest region. Network nodes. Network edge. Benign nodes	–
Aditi Gupta and Rishabh Kaushal	Detecting fake user accounts in Facebook	Data mining	79%

4 Techniques Used in Literature Survey

4.1 Support Vector Machine–Neural Networks

SVM-NN is applied to maximize classification accuracy because it achieves maximum accuracy using a reduced number of features and is implemented on the provided dataset by performing feature reduction by splitting of data testing and training data using eight cross-folds. Neural networks are created by developing neurons and forming a model which are trained and used to predict results. The prediction accuracy is counted separately by using formula.

$$\text{Accuracy} = \text{number of detected accounts/total number of accounts} * 100$$

4.2 Random Forest

Random forest is one of a classification algorithms that is unsupervised in nature. The fundamental concept is a selection of random samples from the provided dataset to create a decision tree followed by result prediction from each tree through voting. Finally, select the most voted results as a final prediction result. It is an ensemble method that achieves the highest accuracy as it reduced the overfitted data in sentiment analysis to identify true and Sybil accounts [5].

4.3 Artificial Neural Networks

Artificial neural networks system framework is a computational processing system that incorporates various interconnected computational nodes that work in a distributed manner to accumulate data from the input to optimize the final output. In implementations of ANN, a connection link is an actual number and the output is determined through a nonlinear input function in each neuron. The objective of the ANN is solving problems like a human brain, for example, in image recognition, pictures containing dogs can be detected by examining pictures manually marked as "dogs," "no dogs" or by using the results to recognize dog in other pictures. This occurs without prior knowledge that dogs have hair, ears, or dog-like heads, for example, characteristics are created by examples they identified automatically. ANN is used for classification, pattern recognition, clustering, regression analysis, prediction, social networks, and even in activities that earlier only humans can do like a painting [8].

4.4 Feature Detection

A convolutional neural network has a special architecture in which complex data characteristics are detected feature is referred to as an "interesting" portion of an in general, image is processed as the first-pixel operation has been performed and each pixel is examined to determine whether a function exists on that pixel. When this belongs to a larger algorithm, the algorithm typically only scans the image in the function field. As a prerequisite for integrated function detection, the Gaussian kernel usually smooths the image input to the size display and calculates one or more feature images that are often expressed concerning the local image derivative.

4.5 Feature Extraction

It is a process of reducing random variables from high-dimensional datasets (data with more than 10 dimensions). It is broken down into two more parts that are feature selection which is a simple approach to find subsets of input variables or attributes and feature extraction. Dimensionality reduction along with feature extraction is applied as a preprocessing step using techniques linear discriminant analysis and K-nearest neighbor to reduce. Feature reduction is required to store time and storage [9].

4.6 Classification

Classification is a complex phenomenon introducing the definition of classes according to the characteristics of the image following the selection of features such as color, texture, and multitemporal data. The feature dataset obtained is trained with supervised or unsupervised learning algorithms. Then, various classification techniques like extraction and segmentation are applied to the trained dataset to get appropriate results. At last, the classification technique is applied to all pixels by suing pixel classification or per-field classification. Image classification covers all unique features of an image and an important factor in the digital environment [3].

5 Proposed Methodology

The proposed algorithm utilizes the convolutional neural network for face recognition [10] and has an age classifier [12]. The features detected from CNN are classified and compared with existing data in the data warehouse for the data comparison.

Convolutional neural networks (CNNs) are a category of deep neural networks, most commonly used for visual imaging processing. They consist of neurons that optimize themselves through learning. CNN is also called as space-invariant artificial neural network (SIANN). CNN is used for image recognition, video analysis, image classification, time series forecasting, recommendation systems, natural language processing, and medical. An input image can be taken by each neuron and operated based on numerous ANN. The only significant distinction between current ANNs and CNNs is that they are primarily used inside objects in the field of pattern detection. CNN comprises of three types of layers. The layers are convolutional layers, pooling layers, and fully connected layers. CNN's fundamental functionality can be explained as:

- The pixel values for the image act as the input and form the input layer.
- The convolutive layer determines the output of the neurons and the linear system is intended to apply CNN's "elemental" activation function to the activation output produced by the previous.
- The pooling layer just downsamples the contribution along with the spatial dimensionality and decreases the number of cases, hence reducing computational times.
- The fully connected layers then perform the indistinguishable tasks as generic ANNs and attempt to generate category results from classifying activations
- CNN makes developing network architecture simpler.

5.1 Proposed Algorithm

The approach this system takes for fake profile detection is too limiting each user, to one and only one account on a particular platform. The system utilized facial recognition and age classification using CNN to create a unique facial print id of the account creator to identify him/her. This helps us to uniquely identify the customer by binding the facial print to the account, eliminating the possibility for a user to create a new fake account. The facial data is utilized to identify two different data, i.e., facial features and then the age prediction. This eliminates the possibility of creating new fake accounts by any user.

The input data after face detection and preprocessing are supplied to the two CNNs for face recognition and age classification. The data are processed by the convolutional network by passing through various layers of convolution, ReLU, pooling/sampling, and finally classified through a fully connected layer.

The process to detect a fake profile involves the following steps: data collection, optimization, face detection, face recognition, age classification, profile association, and profile detection. These steps are explained further in the section.

5.2 Data Collection

The data are collected from the user. This is collected from the social media platform and then passed to the system. This step collects two types of data firstly the profile details like name, age, and facial data from the sensors for face recognition. These data are used to process and identify whether the user who is trying to create an account is genuine or not.

5.3 Optimization

The raw data collected in the collection phase are optimized into the format required. The optimization is one of the important steps before processing of the data as it prepares the data and enhances the data so that during the processing of data the algorithm can produce better results.

5.4 *Face Detection*

Detection involves the detection of the facial data from the collected data and identifies the points required to be processed and specific to the face rather than the whole picture. The detection phase involves facial detection by identifying the points that are of use to us and eliminating the rest unnecessary points. This is done through template matching function. It defines a standard template for all the faces and where the different features can be identified independently like eyes, nose, mouth, contour, etc.

5.5 *Face Recognition*

Face recognition [11] is performed using the convolutional neural network. The data are preprocessed before feeding to the convolutional network. CNN utilizes various hidden layers of convolution, ReLU, and pooling. These layers are arranged in some fashion repeatedly to form the network. After passing through the various hidden layers, the output is put to the fully connected layer of the classifier for classification. The data from the output layer are put to comparison by the dataset in the data warehouse for profile detection.

5.6 *Age Classification*

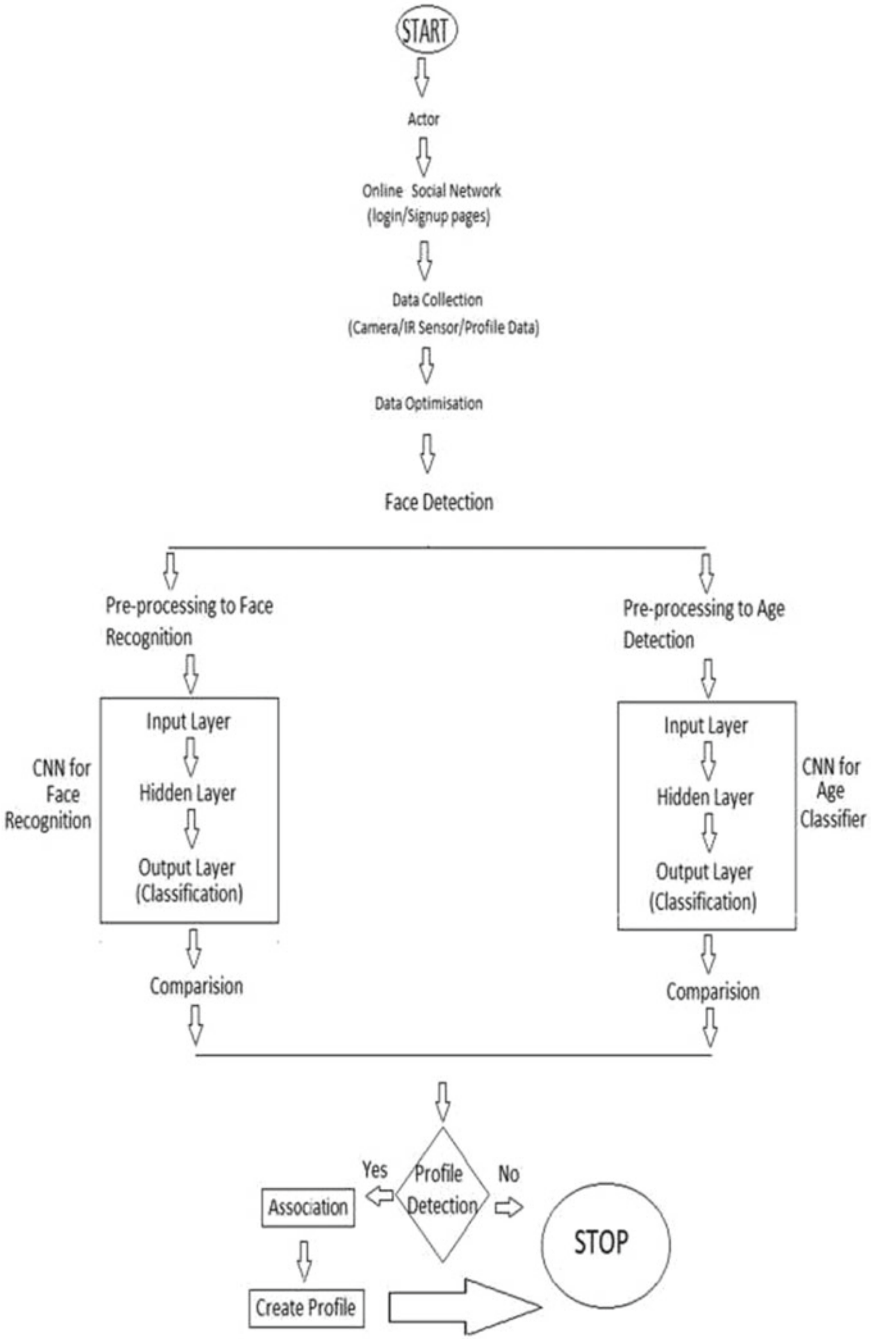
Age classification involves identifying the age of the user using CNN to get near about the age of the user. A different CNN is used for the age classifier. The age predicted by the classifier is given as an estimated range. If the input of the user lies between the range, the profile is allowed for creation. This is an extra parameter just to verify the data input by the user and the verification facial data match the data input by the user at the time of creating the profile [13].

5.7 *Profile Detection*

This involves the main comparison of the data from the data warehouse that utilizes snowflake schema to store the data. The data after the process of classification are passed to this phase. The data are compared with the data in feature data, in the data warehouse. If a match is found, then the profile is flagged as fake and the account is not created. If the face is unique then the age detection data are utilized to compare the age detected by the algorithm with the data entered by the user. If the age matches near about with a minor difference, then the profile is created else it is detected as a fake verification. In the end, if both the tests are accepted then the profile is created and the facial data are moved to the data warehouse for mapping.

5.8 Profile Association

The data are associated with the user profile once it is created. The data are added to the data warehouse creating a unique identification of every use and also the data to which all platforms the user is signed up to are maintained. The data warehouse utilizes are snowflake schema to store the data by identifying the user by its facial patterns. This eliminates the possibility of creating a fake profile of the user on a platform he/she is not utilizing.


START
Actor
Online Social Network
(login/Signup pages)
Data Collection
(Camera/IR Sensor/Profile Data)
Data Optimisation
Face Detection
Pre-processing to Face Recognition
Pre-processing to Age Detection
CNN for Face Recognition
Input Layer
Hidden Layer
Output Layer (Classification)
CNN for Age Classifier
Input Layer
Hidden Layer
Output Layer (Classification)
Comparision
Comparision
Profile Detection
Yes
No
Association
Create Profile
STOP

6 Future Scope

This paper proposes a method that utilizes CNN for face recognition and age classification to eliminate profile. Further implementation can be done using CNN and the accuracy and success of the method can be identified. The paper only proposes the idea of limiting each person to have only one profile on a social media platform to prevent the creation of a fake profile by mapping the creator's facial signature and profile data verification of age.

7 Conclusion

This paper tried to solve the identified problem of fake accounts on the online social media platform by the usage of neural networks and user profile. The paper proposes a system that is expected to prevent the creation of fake profiles as compared to the previous system that utilized machine-neural networks that eliminate fake profiles. Our knowledge does not have any such system that has taken this approach for fake profile elimination, as proposed in the system. The accuracy is not acclaimed as it has to experiment in the future implementation of this system.

References

1. Gayathri, A., Radhika, S., Jayalakshmi, S.L.: Detecting fake accounts in media application using machine learning. Special Issue Published in Int. J. Adv. Netw. Appl. (IJANA)
2. Khaled, S., El-Tazi, N., Mokhtar, H.M.: Detecting fake accounts on social media. In: IEEE International Conference on Big Data, vol. 6, pp. 101–110 (2018)
3. Hudson, B., Voter, B.R.: Profile characteristics of fake twitter accounts. Big Data & Society (2016)
4. Yang, Z., et al.: Uncovering social network sybils in the wild. Trans. Knowl. Discovery Data (TKDD) **8**(1) (2014)
5. Mudasir Ahmad wania, B., Agarwala, N., Jabinb, S., Hussainb, S.Z.: Analyzing real and fake users in Facebook network based on emotions. In: 11th International Conference of Communication System & Networks (2018)
6. Van der Walt, E., Eloff, J.: Using machine learning to detect fake identities: bots vs humans. IEEE Access (2018)
7. Gupta, A., Kaushal, R.: Towards detecting fake user accounts in Facebook. Indira Gandhi Delhi Technical University for Women, Delhi, India
8. Tiwari, V.: Analysis and detection of fake profile over social network. In: International Conference on Computing, Communication and Automation (2017)
9. Şimşek, M., Yilmaz, O., Kahriman, A.H., Sabah, L.: Detecting fake Twitter accounts with using artificial neural networks. In: 2018 Artificial Intelligence Studies (2018); David, I., Siordia, O.S., Moctezuma, D.: Features combination for the detection of malicious Twitter accounts. In: IEEE International Autumn Meeting on Power, Electronics and Computing (2016)
10. Oloyede, M.O., Hancke, G.P., Myburgh, H.C.: Improving face recognition systems using a new image enhancement technique, hybrid features and the convolutional neural network. IEEE Access **6**, 75181 – 75191. https://doi.org/10.1109/ACCESS.2018.2883748

11. Yin, X., Liu, X.: Multi-task convolutional neural network for pose-invariant face recognition. IEEE Trans. Image Process. **27**(2), 964–975. https://doi.org/10.1109/TIP.2017.2765830
12. Rafique, I., Hamid, A., Naseer, S., Asad, M., Awais, M.: Yasir, T.: Age and gender prediction using deep convolutional neural networks. In: 2019 International Conference on Innovative Computing (ICIC). https://doi.org/10.1109/ICIC48496.2019.8966704
13. Chen, S., Zhang, C., Dong, M., Le, J., Rao, M.: Using ranking-CNN for age estimation. in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2017.86