Design And Implement Person Identification Using Python

Shreyash Nitin Kawane
Department of Electronics & Telecommunication
Pune Institute of Computer Technology

Subodh Misal
Department of Electronics & Telecommunication
Pune Institute of Computer Technology

***Abstract*— This technical Paper Describes the design and implementation of a Person Identification using Python. In this Project, Person Identification is highly desirable in applications such as Security Monitoring, Authentication, Criminal Investigation, etc. There are different ways of doing Person Identification using Face Recognition, Voice Recognition, Fingerprint Recognition and Iris Recognition system. But, For this Particular Project we are using Face Recognition and Voice Recognition system.**

Keywords— Face Recognition

I. INTRODUCTION

We are implementing Person Identification using python in this project. For this Implementation we are using Face Recognition and Voice Recognition System. Face Recognition is the ability to identify people according to their Facial Characteristics. Face Recognition is becoming popular for recognizing face of human and it also has become a popular area of research in Computer Vision. Generally, Face Recognition Commonly includes Feature Extraction, Feature Reduction and Recognition or Classification. Face Recognition system can be used to identify people in photos, videos, or in real time. Face recognition uses Computer algorithm to pick out specific, distinctive details about a person's face. These details, such as distance between the eyes or shape of the Chin, are then converted into a mathematical representation and compared to data on other faces collected in Face recognition database. Some face recognition systems, instead of positively identifying an unknown person, are designed to calculate a probability match score between the unknown person and specific face templates stored in the database.

We are also using Voice recognition technique for Person Identification. Voice Recognition is a deep learning technique used to identify, distinguish and authenticate a particular person's voice. It evaluates an individual's unique voice biometrics, including frequency and flow of pitch, and Natural accent. Although, the term "Voice Recognition" and "Speech Recognition" are often used interchangeably, they are distinct, Speech Recognition recognizes Spoken words ; whereas Voice Recognition recognizes Speaker. Voice Recognition is a deep learning technique used to identify, distinguish, and authenticate a particular person's voice. It evaluates an individual's unique voice biometrics, including frequency and flow of pitch and natural accent. Voice Identification uses biological Characteristics of a person's voice to create a voiceprint that is unique to that person. Voice Recognition is most often used as a security measure to confirm the identity of a speaker. Voice Recognition is a contactless, Software-based technology, making it one of the most convenient and readily accepted types of Biometrics, and it is commonly paired with Facial Recognition for Higher levels of Security. It is increasingly utilized for user verification on mobile applications and devices. ''Speech Recognition'' and ''Voice Recognition'' are used interchangeably, they are distinct; Speech Recognition recognizes only spoken Words; whereas Voice recognition recognizes person's unique voice.Although, the term "Voice Recognition" and "Speech Recognition" are often used interchangeably, they are distinct, Speech Recognition recognizes Spoken words ; whereas Voice Recognition recognizes Speaker.

1) Face Recognition System: Face Recognition is a method of identifying or verifying the identity of an individual using their face. Face recognition system can be used to identify people in photos, videos, or in real-time. Face Recognition uses Computer algorithms to pick out specific, distinctive details about a person's face. These details, such as distance between the eyes or shape of the Chin, are then converted into a mathematical representation and compared to data on other faces collected in Face recognition database. Some face recognition systems, instead of positively identifying an unknown person, are designed to calculate a probability match score between the unknown person and specific face templates stored in the database.

2) Voice Recognition System: Voice Recognition is a deep learning technique used to identify, distinguish, and authenticate a particular person's voice. It evaluates an individual's unique voice biometrics, including frequency and flow of pitch and natural accent. Voice Identification uses biological Characteristics of a person's voice to create a voiceprint that is unique to that person. Voice Recognition is most often used as a security measure to confirm the identity of a speaker. Voice Recognition is a contactless, Software-based technology, making it one of the most convenient and readily accepted types of Biometrics, and it is commonly paired with Facial Recognition for Higher levels of Security. It is increasingly utilized for user verification on mobile applications and devices. ''Speech Recognition'' and ''Voice Recognition'' are used interchangeably, they are distinct; Speech Recognition recognizes only spoken Words; whereas Voice recognition recognizes person's unique voice.

3) Face Recognition Algorithms: There are Different types of algorithms that can be used for Face recognition, Some of them are listed below

i) Haar Cascade: Haar Cascades is an object detection method used to locate object on images. The algorithms learn from a large number of positive and negative samples – The former contains an object of Interest and the latter contains anything other than the Object that you are looking for.

i) Convolutional Neural Network (CNN): Convolutional Neural Network (CNN) is one of the breakthroughs of Artificial Neural Networks (ANN) and AI development. It is one of the most popular algorithms in deep learning, a type of machine learning in which model learns to perform classification tasks directly on image, video and Text or sound. The model Shows impressive results in several fields including Computer Vision, Natural Language Processing, and the largest Image classification Dataset.
4) Voice Recognition Algorithms: There are different types of Algorithms that can be used for Voice Recognition in Person Identification which are as follows

i) Pitch Detection algorithm (PDA): Pitch Detection algorithm (PDA) is an algorithm designed to estimate the pitch or fundamental frequency of a oscillating signal using digital recording of a speech.
ii) Gaussian Mixture Model (GMM): GMM is another statistical model that is commonly used for Voice recognition.

II. RELATED WORK

W. S. Zheng [5] suggested a solution in terms of partial person re-identification. They aimed to match partial image with the gallery of full body image. They only focused on the matching between body parts, which are not occluded and fullbody parts. Some of the critical problems are still need to be solved like the output of the occluded person detection model. This case arises when a person standing behind other person and in person detector model, it will detect both partial image and consider both partial image as a single image. The work by Cheng et.al.[6], have done it manually and it seems quite unrealistic in practice. Another suggested approach is patch based matching method. It works to some extent but it need a large amount of calculations. Some of these problems were solved. Jiaxuan Zhuo et.al. [1], proposed a Convolutional Neural Network which differs from the approach of Cheng's solution as they directly compute the matching between occluded person images and full-body person images and propose an Attention Framework of Person Body(AFPB) framework that automatically focuses on the person body by watching various occluded person data generated by an Occlusion Simulator. The Attention Framework of Person Body(AFPB) includes two main components, i.e. Occlusion Simulator (OS) and multi-task losses. The Occlusion Simulator (OS) aims to generate artificial occluded person data. These data are used to simulate occlusion cases using full-body person data The Convolutional Neural Network architecture explored by Zhuo et.al.[6] perform best on ResNet-50 as a backbone architecture. It surpasses the entire previously presented model. Most frequently used datasets in person re-identification problem are CUHK and Market 1501., Person Re-identification task is approached by one shot learning mechanism and Siamese network always give promising result in case of one shot learning. Liang Zheng et.al. [6] proposed very clear and explained survey of re-identification task. Most of the work in person re-identification is done in collaboration of Liang Zheng. His major contribution in reidentification task is making market-1501 dataset. He continuously maintain and upgrade market-1501 dataset. He also supported this research field by maintaining record of state of the art performances and creating new evaluation methods. Typical Person Re-identification (re-occlusion. The second step learns metrics or subspaces for better matching such that distances of the same class are closer than those of the different ones. Recently, with the development of deep learning, there are three kinds of network frameworks applied in person re-id, i.e., classification networks , to obtain 3D models. They reported enhanced performance on the positive and negative samples. In person reidentification, Some hybrid version of deep convolutional Neural Networks is also used. Several end-to-end deep Siamese convolutional neural network architectures have been proposed for human re-identification. These comparisons were done only at final level of architecture. X. zangh et al. [5] proposed a Siamese Long Short Term Memory (LSTM) architecture for human re-identification. Xiao et.al [9], they also make use of LSTM architecture. They tried to increase accuracy by transferring the information to person re-identification task. They trained identity classification and attribute recognition from deep convolutional neural network to learn person information. They extended the architecture of LSTM by a special gate. This Siamese networks [8] are a neural network architecture. In general, a convolutional neural networks loss function is defined such that it learns the similarities and patterns in the images but in Siamese network, instead learning to classify its inputs, the neural network.

III PERSON IDENTIFICATION CLASSIFIER AND ALGORITHMS

i) Haar Cascade:

Here we introduce the Haar classifiers, which were also employed in the first real-time face detector. A machine learning algorithm called a Haar classifier or a Haar cascade classifier finds objects in images and videos. The cascade classifier is composed of a number of stages, each of which contains a group of weak learners. Boosting is used to train weak learners, resulting in a highly accurate classifier from the average prediction of all weak learners. Depending on this forecast, the classifier either chooses to continue on to the subsequent region or decides to notify that an object was discovered (positive) (negative). Because the bulk of the windows do not contain anything of interest, stages are created to reject negative samples as quickly as feasible. Because classifying an object as a non-object would significantly hurt your object detection system, it's crucial to maximise a low false negative rate. You may watch a demonstration of Haar cascades below. "Progress" from the test is shown by the red boxes.

ii) Convolutional Neural Network (CNN):

Convolutional Neural Network (CNN) is one of the breakthrough of Artificial Neural Networks (ANN) and AI development. It is one of the most popular algorithms in deep learning, a type of machine learning in which model learns to perform classification tasks directly on image, video and text or sound. The model shows impressive results in several fields including Computer Vision, Natural Language Processing and the largest Image classification Dataset.

iii) Principal Component Analysis (PCA):

Principal Component Analysis is an unsupervised, non-parametric statistical technique primarily used to dimensionality reduction in machine learning. PCA is the way of reducing the dimensions of the large dataset by transforming it into a smaller dataset contains more information than the lager dataset. By reducing the dataset, we are also reducing the accuracy. However,

PCA works on the principle of trading little accuracy for simplicity. This is because smaller datasets are easier to explore and visualize, thus making Data analysis easier and faster for Machine learning algorithms. Eigen values and Eigen vectors are linear algebra concepts that are used to compute the covariance matrix that determine Principal Component of Data.

iv) Pitch Detection Algorithms (PDA):

PDA is an Algorithm designed to estimate the pitch or Fundamental frequency of a oscillating signal using digital recording of a speech.

v) Teachable Machine IDE:

Teachable machine is a online IDE for voice and face recognition problems. It uses different samples of face and voices for its recognition and identification. It uses transfer learning technique for using training of data and then uploading its model.

vi) MFCC: Mel- Frequency Cepstral Coefficients:

The frequency distribution throughout the window size is summarized by the MFCC. Therefore, it is feasible to examine the sound's frequency and time properties. We locate features for categorization using this audio representation. As a result, it will attempt to transform audio into features based on temporal and frequency properties that will aid in classification. You can watch this movie and read this springer study paper to learn more about MFCC. We will first apply MFCC to a single audio file that we are already using in order to show how it is used in practice.

III. DATASETS AND IMPLEMEMTATION

i) Firstly we have to create dataset for our facial and voice recognition system.We need to create different angles of faces and different samples of voices for out dataset which we can take during the running of our project . These are the sample images we will use for the datatsets. We can record upto minimum of 200 sample for training of data.

ii)Importing Libraries:

We import the required libraries for our voice and face recognition system in the second module. Teacable machine is a significant IDE having different outstanding library that facilitates face and voice analysis. To install the library, just use the IDE or use cmd and use Pip . It offers the necessary building components to create a information retrieval paradigm. TensorFlow is a fantastic package that we will also use for deep learning modelling, so I hope everyone has already loaded it.

iii) Train Model:

Train your model, then instantly test it out to see whether it can correctly classify new examples. Here we can compare the two dataset images we have created in the training mode. We need to train the data set first to export the dataset for further identification. Here we need to train the model of the datasets in the train model option and the two datasets gets the properties verified in this model.We can verify different angles and factors from the datasets like face , eyes , angles ,and also voice frequency etc. Here for voice we need to get minimum of 20 sec samples as datasets for training of the model.Then we need to get background noise as the dataset for training of the model.

iv) Audio Classification for Model Creation:

Here the trained dataset and the samples of audio and images are created as model for extraction. Then the created model needs to be exported for uploading in the code that we have written.

The code then verifies the images and audio trained data and sets us with the result. Now the model generates a link of the particular trained data .Each trained dataset for audio and image generates a link which need to be pasted in the code for recognition of the particular aspect in person identification. Then we need to run the code and paste the links in it.

b) Results and Analysis:

The project report presents the results of a person identification system using Mel-frequency cepstral coefficients (MFCCs) and Teachable Learning IDE. The trained model achieved an accuracy of 97% on the test dataset, demonstrating its effectiveness in accurately categorizing images and voice samples from voice and face images . The evaluation metrics, including precision, recall, and F1-score, consistently showed high values across different emotion classes. The combination proved successful in capturing the nuanced variations and patterns associated with different persons and their identification states. The study highlights the potential of this approach in areas such as affective computing and human-computer interaction. Further enhancements and optimizations can improve the system's accuracy and robustness, leading to a deeper understanding.

CONCLUSIONS AND FUTURE SCOPE

We had learnt how to implement person identification using python. We also Understand the different methods that we had used for implementing this project including Face Recognition, Voice Recognition. In this project we are using python and java as well as teachabale machine Because it is easy to implement and it is having easy syntax.

Now-a-Days, Person identification is become much important as, it is used in our daily life in many fields such as Facial recognition can be used for Criminal Identification in police station, Fingerprints are also another source for doing person identification.

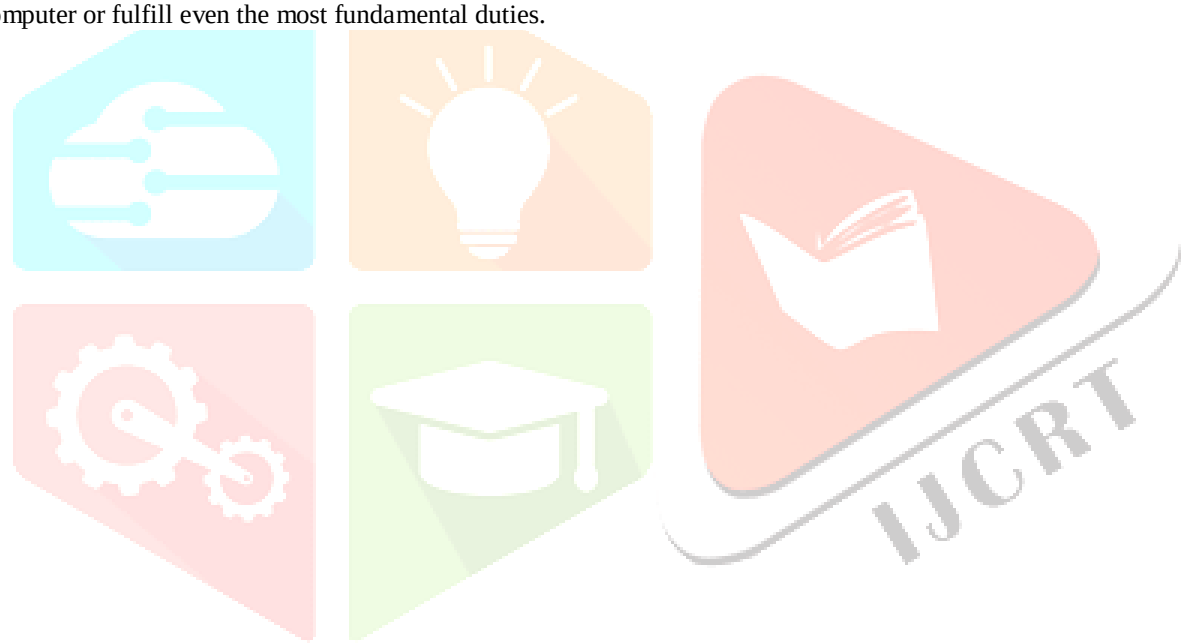
In this project, we also learnt the different algorithms that we used for doing Person identification that helps us in implementing this project.

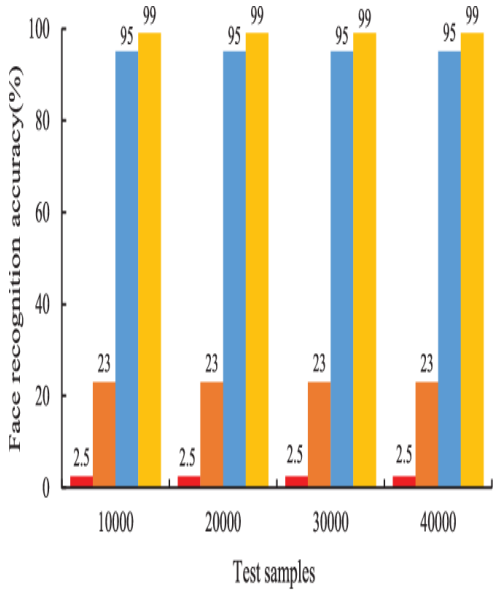
Along with that we learnt to recognition of voice .It can beused in security purposes as well as for maintaining the authenticity of the persons private space .This can also be helpful for company security as well as bank etc.

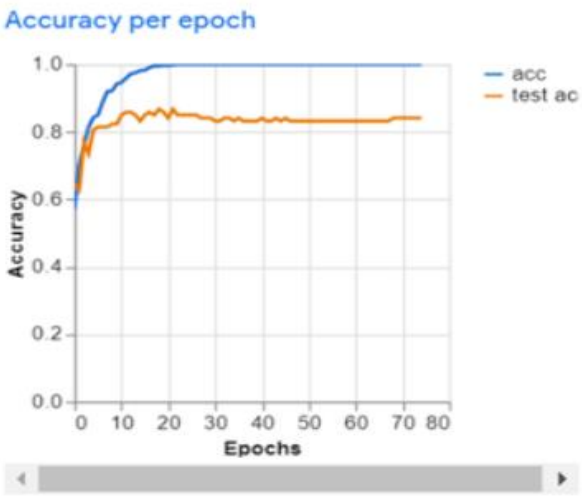
Person Identification is rapidly evolving field, and with the increasing availability of data and advances in machine learning and computer vision, we can expect to see many exciting developments in the coming years. Python is already a widely used language in the field of computer vision and machine learning, and it is likely to remain so in the future. Python provides a wide range of powerful libraries and frameworks for image and video processing, such as OPENCV, sci-kit image and Tensor flow, which can be used for person Identification tasks. One area where we can see a significant progress is in Facial Recognition technology. Facial recognition can already be used to identify individuals in images and videos, and we expect these algorithms to become even more accurate and reliable in the future.

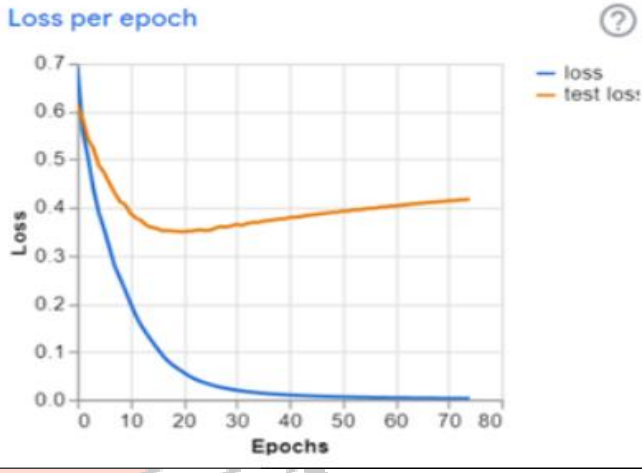
Additionally, we can expect to see more widespread adoption of facial recognition technology for security and surveillance purposes, as well as for applications in the marketing and advertising. In the near Future, Face recognition Technology will likely become m ore ubiquitous. It may be used to track individual's movement out in the world like automated license plate readers, track vehicles by plate numbers.

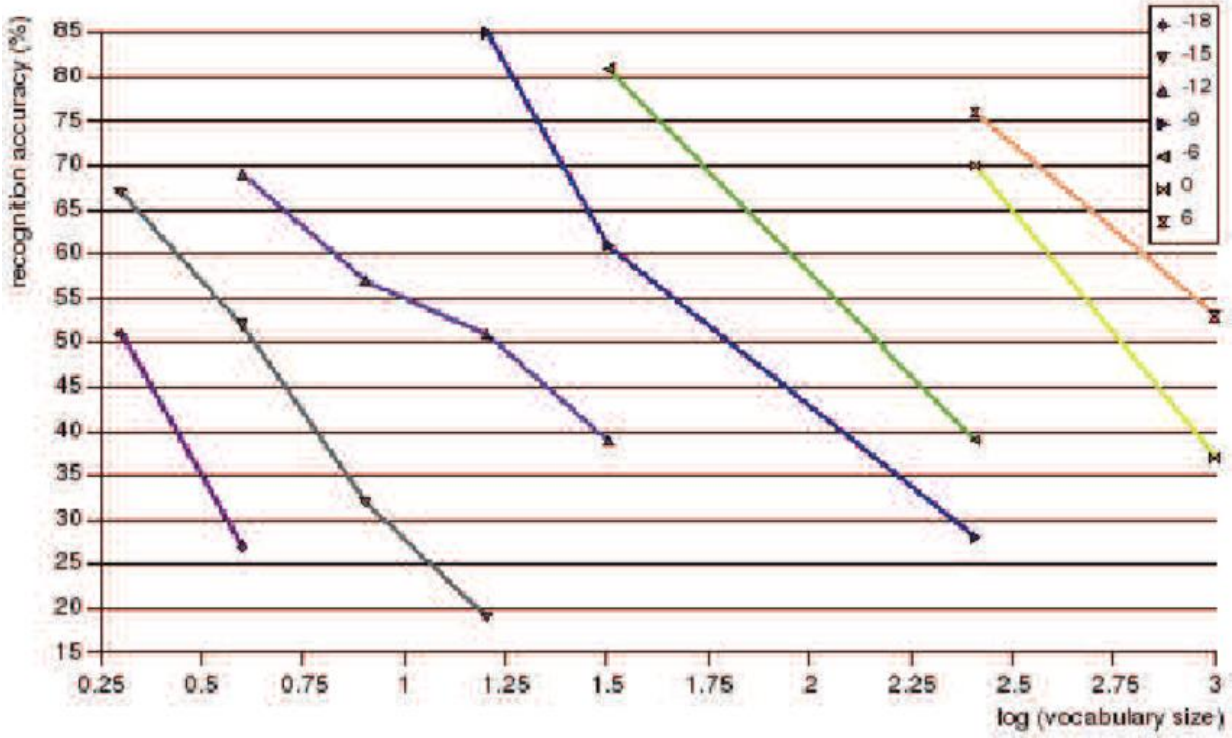
When we speak, the words and messages we relay are less prone to mistakes then when we write or type something out. It's also been a challenge for recognition systems to understand those words clearly.As voice recognition improves the understanding of what we say and how we say it, the accuracy increases and along with that knowledge comes a higher number of tasks the software is capable of completing.In other words, the more the software learns, the better it can adapt to genuine human speech patterns and inflection and the easier it will be for it to carry out more advanced commands.This means that instead of menial tasks like the aforementioned light switching or music playing, you can trust your digital assistant to take accurate dictation, use auditory notes instead of handwritten ones, and apply the tech to areas with higher stakes than your home (like commercial, medical, or industrial applications). Recognition software though really finds its purpose through those with disabilities who might otherwise be unable to use a computer or fulfill even the most fundamental duties.




Face recognition accuracy(%)
Test samples
10000
20000
30000
40000
100000
200000
300000
400000


Accuracy per epoch
acc
test ac
Accuracy
Epochs


Loss per epoch
loss
Loss
Epochs


recognition accuracy (%)
log (vocabulary size)

REFERENCES

[1] International journal of advance Computer Science and Application: Study of Face Recognition Techniques, Vol 9, Nov 2018.

[2] Jain A.K. Technology: Biometric Recognition. Nature 449, 38-40 (2001)

[3] Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. ACM Comput. Surv. 35, 399–458 (2003)

[4] D. A. Reynolds, "An overview of automatic speaker recognition technology", *2002 IEEE International Conference on Acoustics Speech and Signal Processing*, vol. 4, pp. IV-4072, 2002.

[5] S. Paulose, D. Mathew and A. Thomas, "Performance evaluation of different modeling methods and classifiers with MFCC and IHC features for speaker recognition", *Procedia computer science*, vol. 115, pp. 55-62, 2017.

[6] S. K. Wong and S. M. Yiu, "Location spoofing attack detection with pre-installed sensors in mobile devices," J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl, vol. 11, no. 4, pp. 16–30, 2020.

[7] A. S. Kitana, T. Issa, and W. G. Isaac, "Towards an epidemic SMS-based cellular botnet," J. Internet Serv. Inf. Secur, vol. 10, no. 4, pp. 38–58, 2020

[8] "Global Social media Attacks Rate Among Businesses," 2019.

[9] "5 Shocking Insights into the Social Network Habits of Security Professionals," 2017, https://thycotic.com/company/blog/2017/05/30/5-shocking-insights-into-the-social-network-habits-of-securityprofessionals-and-infographic/.

[10] G. S. Kasturi, A. Jain, and J. D. Singh, "Detection and classification of radio frequency jamming attacks using machine learning," J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl, vol. 11, no. 4, pp. 49–62, 2020.

[11] A. L. Marra, F. Martinelli, F. Mercaldo, A. Saracino, and M. Sheikhalishahi, "A distributed framework for collaborative and dynamic analysis of android malware," J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl, vol. 11, no. 3, pp. 1–28, 2020.

[12] . Berbecaru, A. Lioy, and C. Cameroni, "Supporting Authorize-then-Authenticate for Wi-Fi access based on an electronic identity infrastructure," J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl, vol. 11, no. 2, pp. 34–54, 2020.

[13] S. H. K. Wong and S. M. Yiu, "Identification of device motion status via Bluetooth discovery," J. Internet Serv. Inf. Secur, vol. 10, no. 4, pp. 59–69, 2020

[14] J. A. Unar, W. C. H. Seng, and A. Abbasi, "A review of biometric technology along with trends and prospects," Pattern Recognition, vol. 47, no. 8, pp. 2673–2688, 2014.

[15] Y. Lu, L. Li, H. Peng, and Y. Yang, "An enhanced biometric-based authentication scheme for telecare medicine information systems using elliptic curve cryptosystem," Journal of Medical Systems, vol. 39, no. 3, pp. 32–38, 2015.

[16] S. Chatterjee, S. Roy, A. K. Das, S. Chattopadhyay, N. Kumar, and A. V. Vasilakos, "Secure biometric-based authentication scheme using Chebyshev chaotic map for the multi-server environment," IEEE Transactions on Dependable and Secure Computing, vol. 15, no. 5, pp. 824–839, 2018.

[17] P. Padma and S. Srinivasan, "A survey on biometric-based authentication in cloud computing," in Proceedings of the International Conference on Inventive Computation Technologies, pp. 1–5, ICICT), Coimbatore, India, August 2016.

[18] J. Yang and R. K. Das, "Long-term high-frequency features for synthetic speech detection," Digital Signal Processing, vol. 97, Article ID 102622, 2020.