

## AN ENHANCED MULTIMODAL BIOMETRIC SYSTEM BASED ON CONVOLUTIONAL NEURAL NETWORK

LAWRENCE OMOTOSHO<sup>1</sup>, IBRAHIM OGUNDOYIN<sup>1</sup>, OLAJIDE ADEBAYO<sup>1</sup>,  
JOSHUA OYENIYI<sup>1</sup>

<sup>1</sup>*Department of Information and Communication Technology, Osun State University,  
Osogbo, Osun State, Nigeria*

**Abstract:** Multimodal biometric system combines more than one biometric modality into a single method in order, to overcome the limitations of unimodal biometrics system. In multimodal biometrics system, the utilization of different algorithms for feature extraction, fusion at feature level and classification often to complexity and make fused biometrics features larger in dimensions. In this paper, we developed a face-iris multimodal biometric recognition system based on convolutional neural network for feature extraction, fusion at feature level, training and matching to reduce dimensionality, error rate and improve the recognition accuracy suitable for an access control. Convolutional Neural Network is based on deep supervised learning model and was employed for training, classification, and testing of the system. The images are preprocessed to a standard normalization and then flow into couples of convolutional layers. The developed multimodal biometrics system was evaluated on a dataset of 700 iris and facial images, the training database contain 600 iris and face images, 100 iris and face images were used for testing. Experimental result shows that at the learning rate of 0.0001, the multimodal system has a performance recognition accuracy (RA) of 98.33% and equal error rate (ERR) of 0.0006%.

**Keywords:** multimodal, biometric system, convolutional neural network

### 1. INTRODUCTION

A biometric system is an identification system that can analyze unique physiological or behavioral traits of an individual for authentication purpose. Biometrics traits are always unique, measurable, permanent or automatically validated [1]. Biometric systems using a single biometric trait is called unimodal biometric system while a multimodal biometric system combines more than one biometric traits or for authentication purpose.

Unimodal biometric system suffers from different problems such as lack of uniqueness, restricted degrees of freedom, non-universality, intra-class variation, noisy data, vulnerable to spoofing, and unacceptable error recognition rates [2]. Unimodal biometric systems have limitations in terms of accuracy, enrolment rates, and susceptibility to spoofing [3]. Multimodal biometric system combines more than one biometric trait into a single identification system in order to overcome the limitations of unimodal biometrics system and improve recognition accuracy. The use of several entities in multimodal biometric system makes it more reliable, secure, accurate and robust.

In recent years, convolutional Neural Network (CNN) has become a mainstream in pattern and image recognition due to the uniqueness in its architecture. It has the inherent capacity to perform segmentation, feature extraction

---

\* Corresponding author, email: [thejoshoyeniya@gmail.com](mailto:thejoshoyeniya@gmail.com)

and classification in one module, which has widely been used in a variety of area. It has increasingly been used in computer vision and has achieved measurable success in image and video recognition [4]. CNN is a feed-forward network with the ability to extract features from the input images, and then classifying the extracted feature [5].

Multimodal biometric recognition system has been adopted in various applications such as information security, banking, access control etc. The fusion of data can be carried out at different levels in a multimodal biometric system, the fusion can occur at the sensor level, feature level, classifier level and rank level. Generally, biometrics systems are classified in static and dynamic biometric systems [6].

Convolutional neural network (CNN) has been recognized as an essential feature learner and classifier in various object recognition, object detection and classification methods. Convolutional neural networks have been mostly used as classifiers, but they are also efficient tools to extract and represent discriminative features from the raw data at different levels of abstraction [7]. In comparison to hand-crafted features, the use of CNN as domain feature extractor has proven to be more promising when facing different modalities such as face, iris and fingerprint [8].

In multimodal biometrics system, the use of different algorithms for feature extraction, fusion at feature level and classification often to complexity and make fused biometrics features larger in dimensions. In this paper, we developed a face-iris multimodal biometric recognition system based on convolutional neural network for feature extraction, fusion at feature level, training and matching to reduce dimensionality, error rate and improve the recognition accuracy suitable for an access control. Convolutional Neural Network is based on deep supervised learning model and was employed for training, classification and testing of the system.

Some authors [9] proposed a multimodal biometric human recognition system based on fuzzy vault fusion. In their work, ear and fingerprint modalities were used for personal authentication and implemented in MATLAB. The phases of the proposed recognition scheme were pre-processing; extraction of features; development of clustered feature vectors; and merger and recognition. A better accuracy value of 98.8166 % on average for the identification of individuals with the fingerprint and ear modalities was enabled by the results of their work. False acceptance rate, false rejection rate and genuine acceptance were the assessment measures used.

A new method for extracting features in spatial domain from palmprints and iris was proposed by [10]. Thepade's Block Truncation Coding level 2 was used to reduce the feature vector size and to improve accuracy in form of genuine acceptance rate (GAR) in multimodal biometric identification they took the iris and palmprint together at matching score level. They also considered color spaces on iris images for improvement in genuine acceptance ratio (GAR). The extraction method improves the accuracy and reduces the error rate. The accuracy of the test was based on genuine acceptance rate.

A bimodal biometric student attendance system using fingerprint and face traits was designed by [11]. The system make use of face and fingerprint trait to take students attendance, they capture the students' faces using webcam and preprocessed the captured faces by converting the color images to grey scale images. The fingerprint and facial images of each user were stored along with their particulars in a database. The system had a recognition accuracy of 87.83%. They were able to explore the use of bimodal biometrics to improve the recognition accuracy of automated student system.

The authors [12] integrated fingerprint and face biometric to improve performance in an access control system. They considered restoration of distorted and misaligned fingerprints caused by environmental noise such as oil, wrinkles, dry skin, dirt, displacement etc. Hybrid Modified Gabor Filter-Hierarchical Structure Check (MGF-HSC) system model was employed to optimize the noisy, distorted, and misaligned fingerprint. Fast Principal Component Analysis (FPCA) algorithm was employed to address the problem of different face conditions (face distortions) such as lighting, blurriness, pose and head orientation. The algorithms employed improved the quality of distorted and misaligned fingerprint image. They were able to improve the recognition accuracy of distorted face during authentication. Experimental result of the system gives 97.86% recognition accuracy.

Also, [13] designed a novel multimodal recognition system that trains a deep learning network to automatically learn features after extracting multiple biometric modalities from a single data source, i.e., facial video clips were presented. In the facial video images, they use various modalities such as left ear, left profile face, frontal face, right profile face, and right ear, and train supervised denoising auto encoders to automatically extract robust and non-redundant characteristics. The learned features were used to train modality specific sparse classifiers in order to perform multimodal recognition. The restricted facial video dataset and unconstrained facial video dataset

experiments resulted in 99.17% and 97.14% rank-1 recognition scores, respectively. The precision of multimodal recognition revealed the superiority and robustness of the proposed solution regardless of the lighting, non-planar motion, and variations in the video clips.

Other study [14] proposed a scheme of multimodal biometric system based on texture information extracted from face and two iris (left and right) using hybrid level of fusion. Face and iris were biometrics. The proposed schemes were tested on CASIA Iris Distance database for verification. The experimental results show that the proposed multi-modal biometric system yields attractive performances of up to 0.24% in terms of Equal error rate, false acceptance rate of 0.06% Genuine acceptance rate of 99.5%.

Also, [15] presented a novel method for access control system using ear and tongue biometric modalities. The images were obtained using a digital camera, fusion was carried out at the feature extraction level. The ear and tongues features were extracted using Principal Component Analysis (PCA) was and Self Organizing Feature Map Neural Network (SOFM) was used for training and testing of the system. The method was evaluated using 5000 ear images and tongue images. The performance evaluation metrics employed were false acceptance rate, false rejection rate, equal error ate and performance recognition accuracy. The accuracy and performance of system was measured by plotting False Acceptance Rate (FAR) and False Rejection Rate (FRR). Experimental results show that the method has 99.78% performance recognition and 0.003% equal error rate (ERR), the fusion results of ear and tongue images showed an improved performance and a huge step closer for user access control, the results also revealed that multimodal biometric authentication system is much more reliable and useable in real-time authentication systems.

## 2. EXPERIMENTAL SETUP

The methodology includes image acquisition, preprocessing, feature extraction and matching. Iris and face traits were employed as the biometrics modalities for the multimodal biometric recognition system and simulated on MATLAB. The images were preprocessed before the extraction of features, the extracted features of iris and face were fused, Convolutional neural networks (CNNs) were employed for training, classification and testing of the system.

### 2.1. CNN architecture

The four types of layers for a convolutional neural network are the convolutional layer, the pooling layer, the ReLU correction layer and the fully-connected layer (Figure 1).

**Convolution layer:** This layer detects the presence of a set of features in the images received as input, then condenses the input image by extracting features of interest from the image and produces feature maps, a feature is seen as a filter. At this stage, the input image  $I_{(m)}$  is convolved with the filter kernel  $F_{(k)}$ , the dimensions of the filter kernel used for convolution are 2x2 as shown in (1).

$$L(\text{conv}) = I_{(m)}F_{(k)2 \times 2} \quad (1)$$

where  $L_{(\text{conv})}$  is convolution layer,  $I_{(m)}$  is the input image and  $F_{(k)}$  is the filter kernel.

**Rectified linear unit (ReLU) layer:** This serves as an activation function, it uses a nonlinear to minimize the linearity introduced in the convolutional layer.

$$I \text{ RELU}(x, y) = \max(L(\text{conv}), (x, y), 0) \quad (2)$$

**The pooling layer:** At this layer, data dimensionality are reduced and this improved the efficiency of the network and avoids over-learning. It also minimizes the computation time and also control the over fitting.

**Fully connected layer:** Classification takes place at this layer, it classifies the image as an input to the network: it returns a vector of size N, where N is the number of classes in our image classification problem. Each element of the vector indicates the probability for the input image to belong to a class and determines the relationship between the position of features in the image and a class.

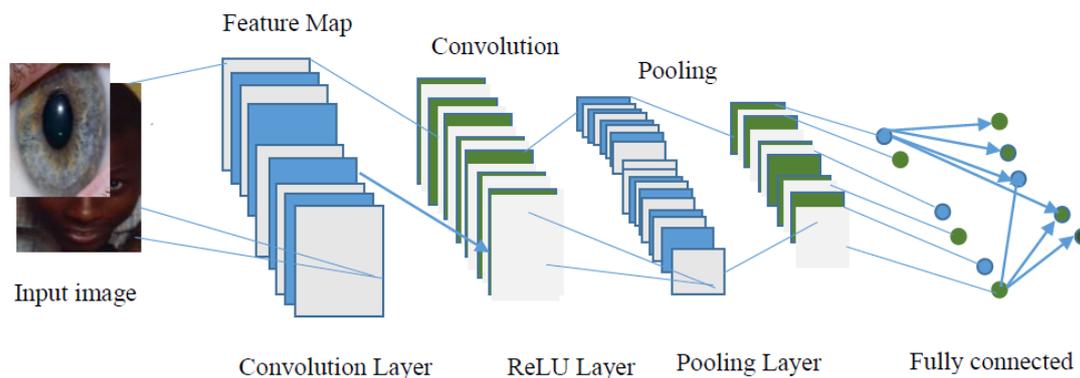


Fig. 1. CNN architecture.

## 2.2. Image acquisition

Image acquisition is the first stage of any pattern recognition process as shown in Figure 2. The biometrics traits used were iris and face. The total number of images used is 700; these images are stored and were used as training and testing datasets.

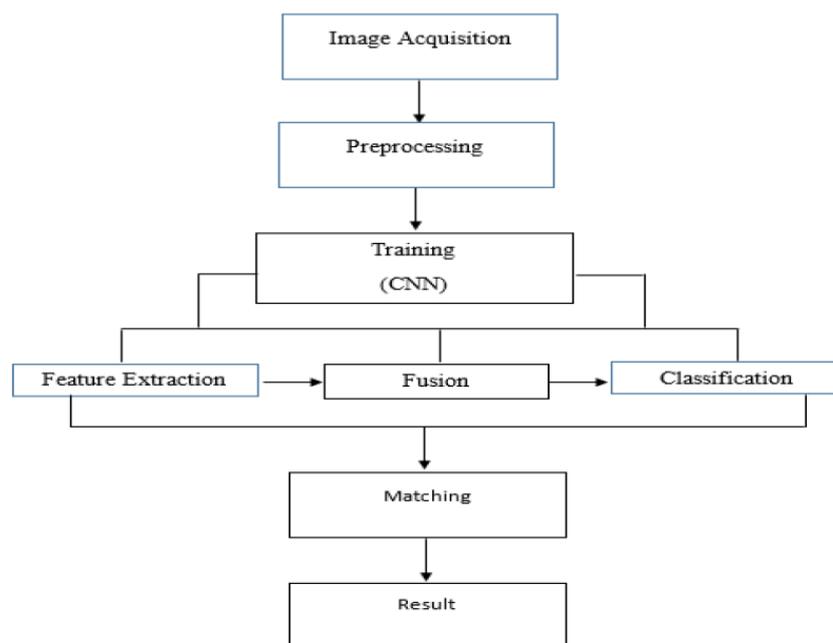


Fig. 2. Block diagram of the multimodal system.

## 2.3. Preprocessing

At the preprocessing stage, region of interest (ROI) was extracted from the acquired images and normalization was done. The purpose of the pre-processing is to reduce or eliminate some of the variations in the images due to illumination and improve the visual quality. To maintain the uniformity in the database, all the images are resized to  $277 \times 277$  pixels by convolutional neural networks (CNNs).

## 2.4. Training

At the training stage, mini batch gradient descent optimization algorithm is use for the learning. In mini-batch gradient descent algorithm (n) number of training dataset samples are divided into small batches (b), then the model coefficients are updated using model error.

$$T = \frac{n}{b} \quad (3)$$

where T represent the total n is number of iteration per training epoch and b represent the batches. The weights (w) of CNN are optimized using error function defined in equation (4) below.

$$\sum_t [f(w)] = \frac{1}{b} \sum_i^{tb} = (t-1)b + 1 f(w, x_0) \quad (4)$$

where  $X_0$  is the sample of training data and W represent the weight. At each iteration, the weights are updated by rule mini batch gradient descent update rule with learning rate given in equation (5).

$$W^{t+1} = W^t - \mu \nabla_w \sum_t [f(w)] \quad (5)$$

#### 2.4.1. Feature extraction

The iris and facial images features were extracted using convolutional neural networks. To extract features from CNN model, the network was trained with the last sigmoid/logistic dense layer. Convolutional neural network is based on deep supervised learning model. The first layer of CNN is convolution layer in which the original image is convolved. The layer is also called hidden feature extractor which describes the internal connectivity of the image region.

#### 2.4.2. Fusion

The fusion of neurons obtained from the third layer of CNN applied to face and iris is performed using concatenation of the two features to increase the discriminant nature of the feature set.

$$Fv = Xo + Yo \quad (6)$$

where  $X_0$  is face feature vector,  $Y_0$  is Iris feature vector and  $F_v$  is the Fusion layer value.

#### 2.4.3. Classification

Alexnet CNN classifier which is a deep convolutional neural network was used for classification.

$$\langle V.X_0 \rangle + b_0 \geq Y_0, = 1 \quad (7)$$

$$\langle V.X_0 \rangle + b_0 \geq -1, Y_0 = 1 \quad (8)$$

where V represent the vector,  $b_0$  is the bias value,  $X_0$  is CNN feature set and  $Y_0$  represent the class label.

### 2.5. Matching

Matching is a method where the feature extracted from the image called user template are compared with the template of the image stored in the database. It helps us to verify the authenticity of the person.

## 3. RESULTS AND DISCUSSION

The simulation was performed with the machine specifications as shown in Table 1 and implemented using MATLAB 2018a. The iris data set were obtained from CASIA database while the face images were captured using webcam. The developed multimodal biometrics system was evaluated on a dataset of 700 iris and facial images. The training database contain 600 iris and face images, 100 iris and face images for testing. The performance of the system was evaluated based on Recognition Accuracy (RA) and Equal error rate (EER). Table 2 shows the parameters used for the simulation and training of the convolutional neural network. Table 3 shows the result for the training at different epoch and iteration at the learning rate of 0.00001. Figures 3, 4 and 5 shows the result of different performance metrics for the system. Experimental result shows that at the learning rate of 0.0001, the multimodal system has a performance recognition accuracy of 98.33% and equal error rate of 0.0006%

At an epoch of 1, the training iteration value is 1 with a mini batch accuracy of 4.69% and mini batch loss of 3.5405. At an epoch of 8, the training iteration value is 50 with a mini batch accuracy of 96.88% and mini batch loss of 0.2321. At an epoch 10 of the training iteration value is 1 with a mini batch accuracy of 96.88% and mini batch loss of 0.1084. There is a significant change in the batch accuracy level and batch loss value as the iteration increases at different epoch.

Table 1. Machine specification for simulation.

PARAMETERS	VALUE
Software	MATLAB R2018a
Machine name	OYENIYI – PC
Operating System	Windows 10, 64-bit
Language	English (Regional Setting: English)
System Manufacturer	HP 250
System Model:	D7329939-131A-46D6-B825-88BFA061B700
Processor	Intel(R) Pentium ® CPU 2020M @ 2.40GHZ, 2.40GHZ
RAM	4GB

Table 2. Parameter for the training.

Parameter	Specification
Learning rate	0.00001
Mini batch size	64
Minimum epoch	10
Filter size	1

Table 3. Result for the training at different epoch.

Epoch	Iteration	Time	Mini batch accuracy	Mini batch loss
1	1	0:00:26	4.694%	3.5405
8	50	0:14:01	96.88%	0.2321
10	70	0:19:08	96.88%	0.1084

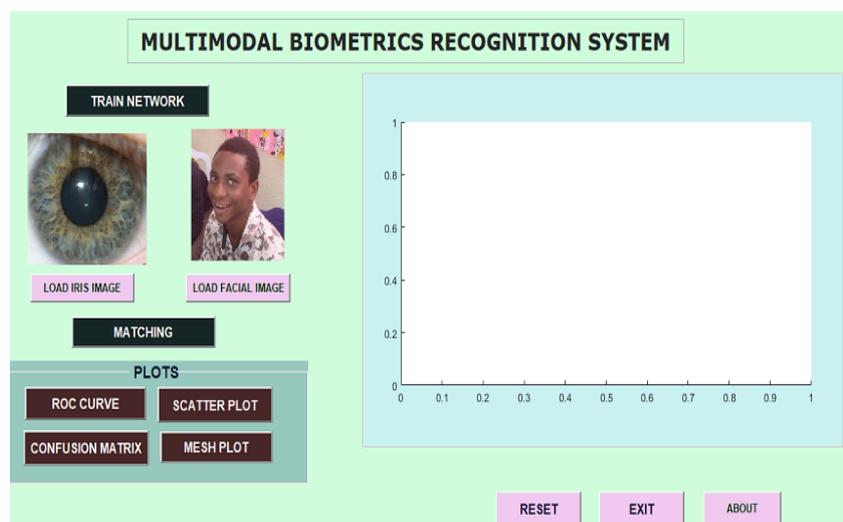


Fig. 3. User interface of the developed multimodal biometrics recognition system.

Figure 3 show the graphical user interface (GUI) of the system. Figure 4 show that there is a considerable correlation between the training and testing datasets. The data plotted as points in the scatter plot shows the pattern in the datasets appears to be correlated during the training and matching. The pattern of the resulting dots reveals that relationship between the two is considerably strong. Thus, the developed multimodal system is reliable for real time authentication.

Figure 5 show the forward and back propagation during the training, forward propagation denote the step where the input data are transformed into output through CNN layers, back propagation propagate the errors from a layer to the previous one and compute the derivate of the error with respect to weight and biases. The developed multimodal biometrics system has a recognition accuracy of 98.33%.

Figure 6 show the matching result, the test data set was subjected to verification for identification of individuals whose biometric traits are stored in the database.

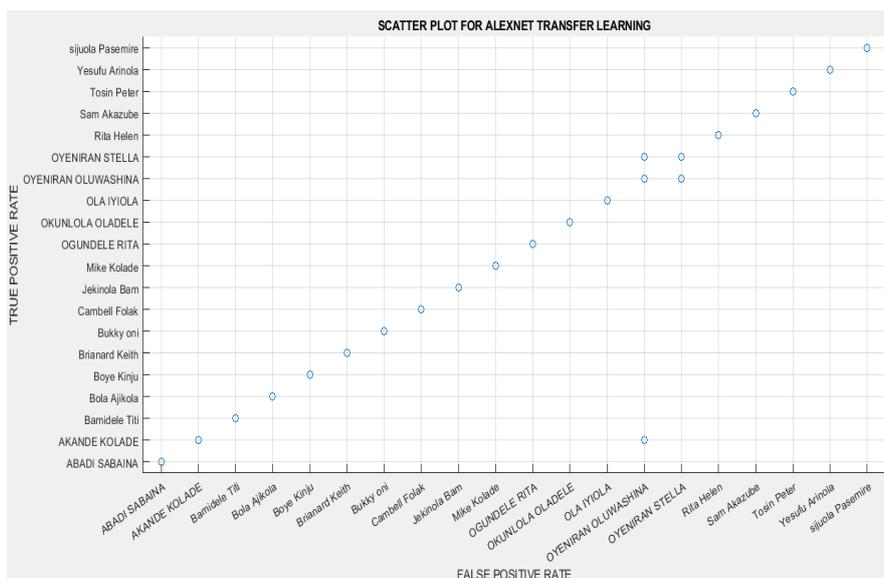


Fig. 4. Scatter plot for the multimodal biometrics system.



Fig. 5. Training of the system using CNN.



Fig. 6. Matching result of the multimodal biometric recognition system.

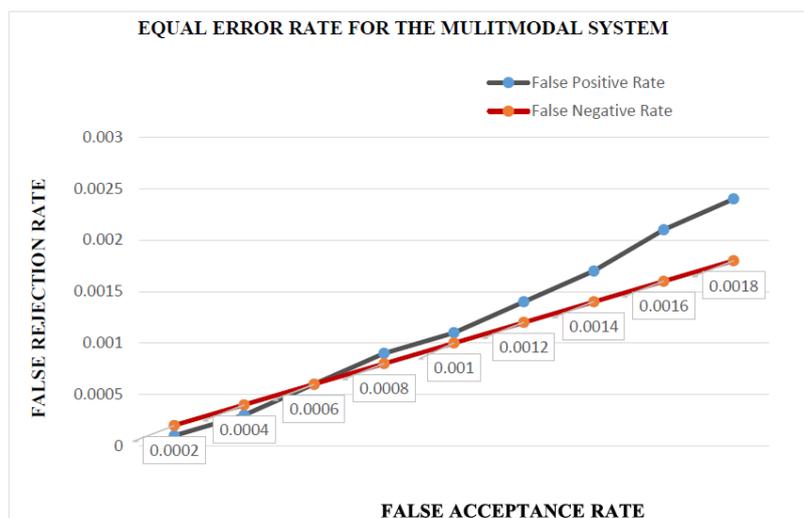


Fig. 7. Equal error rate.

Equal error rate (Figure 7) is used to determine the threshold values for false acceptance rate and false rejection rate. The developed multimodal biometrics system has a recognition accuracy of 98.33% and ERR value of 0.0006%. The result show that the multimodal biometrics system based on convolutional neural networks (CNNs) for feature extraction and classification improves the recognition accuracy and has a lower equal error rate. The proposed method based on convolutional neural networks for both training, classification and testing improved the performance of a biometric authentication with better accuracy. The system achieves an attractive improvement in terms of equal error rate (EER) compared to the unimodal biometrics system and several multimodal systems reviewed in this work. This method based on CNN provides a simplified and efficient method for feature extraction and classification in multimodal biometric system, it also facilitate a suitable structure and reasonable environment for training process feed-forward and backward propagation of errors.

#### 4. CONCLUSION

Multimodal biometric recognition system is much more reliable and effective in real-time authentication systems. Experimental results have demonstrated that the proposed method based on convolutional neural networks for both training, classification and testing enhanced the performance of a biometric authentication with better accuracy and considerable improvement in term of EER which is suitable for an access control system. The use of convolutional neural network for feature extraction, training, and classification and testing reduces complexity and dimensionality in multimodal biometric recognition system and improved the recognition accuracy. Also, it is effective and efficient to be integrated for personal identification. Future studies should consider the impact of such optimization learning algorithms on a multimodal system consisting of different biometric characteristics and experiments with a greater number of subjects in datasets.

#### REFERENCES

- [1] Cherrat, E.M., Alaoui, R., Bouzahir, H., Jenkal, W., High-density salt-and-pepper noise suppression using adaptive dual threshold decision based algorithm in fingerprint images, *Intelligent Systems and Computer Vision*, 2017, p. 1-4.
- [2] Ammour, B., Bouden, T., Boubchir, L., Face-iris multimodal biometric system based on hybrid level fusion, *Biometrics*, vol. 7, no. 5, 2018, p. 482-489.
- [3] Sasidhar, K., Kakulapati, V.L., Ramakrishna, K., Kailasa, R.K., Multimodal biometric systems – study to improve accuracy and performance, *International Journal of Computer Science and Engineering Survey*, vol. 1, no. 2, 2010, p. 54-61.
- [4] Al-Waisy, A.S., Qahwaji, R., Ipson, S., Al-Fahdawi, S., Tarek, A.M., A multi-biometric iris recognition system based on a deep learning approach, *Pattern Analysis and Applications*, vol. 21, no. 3, 2017, p. 783-802.
- [5] Tiong, L., Kim, S.T., Ro, Y., Multimodal face biometrics by using convolutional neural networks, *Journal of Korea Multimedia Society*, vol. 20, no. 2, 2017, p. 170-178.

- [6] Shende, P., Dandwate, Y.H., Convolutional neural network based multimodal biometric human authentication using face, palm veins and fingerprint, *International Journal of Innovative Technology and Exploring*, vol. 9, no. 3, 2020, p. 771 -777.
- [7] Soleymani, B., Dabouei, A., Kazemi, H., Dawson, J., Nasrabadi, N.M., Multi-level feature abstraction from convolutional neural networks for multimodal biometric identification, *24th International Conference on Pattern Recognition*, 2018, p. 3469-3476.
- [8] Madane, M., Thepade, S., Score level fusion based bimodal biometric identification using thepade's sorted n-ary block truncation coding with variod proportions of iris and palmprint traits, *Procedia Computer Science*, vol. 79, 2016, p. 466-473.
- [9] Vinothkanna, R., Wahi, A., Fuzzy vault fusion based multimodal biometric human recognition system with fingerprint and ear, *Journal of Theoretical and Applied Information Technology*, vol. 59, 2014, p. 304-316.
- [10] Soleymani, S., Torfi, A., Dawson, J., Nasrabadi, N.M., Generalized bilinear deep convolutional neural networks for multimodal biometric identification, *IEEE International Conference on Image Processing*, 2018, p. 1-3.
- [11] Atuegwu, C., Kennedy, O., Noma, O.E., A bimodal biometric student attendance system, *IEEE 3rd International Conference on Electro-Technology for National Development*, 2017.
- [12] Zuva, T.O., Esan, A., Ngwira, S.M., Hybridization of bimodal biometrics for access control authentication, *International Journal of Future Computer and Communication*, vol. 3, no. 6, 2014, p. 41-44.
- [13] Sayan, M., Mohamed, A., Shihab, S., Multimodal biometrics recognition from facial video via deep learning, *Signal and Image Processing: An International Journal*, vol. 8, no.1, 2017, p. 67– 75.
- [14] Adebayo, O.Y., Adigun, A.A., Isola, E.O., A ear and tongue based multimodal access control system, *Computer Science Series*, vol. 18, no. 2, 2019, p. 85-92.
- [15] Soleymani, S., Torfi, A., Dawson, J., Nasrabadi, N., Generalized bilinear deep convolutional neural networks for multimodal biometric identification, *IEEE International Conference on Image Processing*, 2018, p. 763-767.