Development of a Multimodal Biometric Security System using Modified Convolutional Neural Network

O.G. Atanda
Department of Cyber-security
LAUTECH, Ogbomoso.
dayo.oladayoo@gmail.com

A.O. Afolabi
Department of Cyber-security
LAUTECH, Ogbomoso.
adefolabius@yahoo.com

A.S. Falohun
Department of Computer Engineering
LAUTECH, Ogbomoso.
asfalohun@lautech.edu.ng

ABSTRACT

Biometrics is the biological measurement of physiological or behavioral attributes of humans. These characteristics are unique to each individual and remain unaltered throughout human lifetime. Several multimodal biometric security systems have been developed using convolutional neural network but few of them have been able to handle the challenges of complexities associated with CNN in terms of recognition rates and time. In this work, a multimodal biometric security system that uses a modified CNN (CNN-GA) for feature extraction and classification was developed.

The System was tested on a database consisting of 1026 trained images and 684 probe images of face, ear and iris biometrics. The design, implementation and running/testing of the entire system were done on MATLAB R2016a programming platform. The multimodal images were first preprocessed. Feature extraction and classification were carried out using Convolution Neural Network-Genetic Algorithm (CNN-GA). It was the optimal classification result that was used to make final decision on whether to accept or reject probe images. The performance evaluation of the developed system was carried out using false positive rate, sensitivity, specificity, precision, recognition accuracy and recognition time.

The result shows that at varying threshold values of 0.20, 0.35, 0.50 and 0.76, the CNN-GA outperforms the standard CNN as applied to the developed system in terms of sensitivity, specificity, precision, recognition accuracy and time. At the threshold value of 0.76, CNN-GA achieved a sensitivity of 97.66%, specificity of 98.25%, Precision of 99.40%, recognition accuracy of 97.81% and recognition time of 455.54Secs while the standard CNN achieved a sensitivity of 95.91%, specificity of 92.98%, Precision of 97.62%, recognition accuracy of 95.18% and recognition time of 565.02Secs.

Key words: Multimodal Biometrics, Weighted Sum Rule, Modified CNN, Receiver Operating Characteristics.

1 INTRODUCTION

Biometrics recognition according to Rahal (2006) refers to the automatic recognition of individuals based on their physiological and behavioral characteristics. These characteristics are unique to each individual and remain unaltered throughout human lifetime. Utilizing biometrics for personal authentication is becoming more accurate than traditional methods (such as the utilization of passwords or Personal Identification Numbers - PINs) and more convenient (nothing to carry or remember). The etymology of biometrics is derived from the greek word "bios", which means life and "metron" which means "to measure", thus biometrics means life measurement (Modi, 2011).

Biometrics is not just about security, it is also about convenience. The need for biometrics can be found in a wide range of commercial, military and other related applications. Biometrics is set to infiltrate nearly all aspects of the economy and our daily lives. The use of biometric technologies involving mathematical analysis of a unique trait such as face, ear, and iris has been adopted worldwide and on large scale.

Most biometric systems deployed in real-world applications are unimodal, i.e., they rely on the evidence of a single source of information for authentication (e.g. single fingerprint or face). Multi-biometric systems are able to meet the stringent performance requirements imposed by various applications. They address the problem of non-universality, since multiple traits ensure sufficient population coverage. They also deter spoofing since it would be difficult for an impostor to spoof multiple biometric traits of a genuine user simultaneously. Furthermore, they can facilitate a challenge response type of mechanism by requesting the user to present a random subset of biometric traits thereby ensuring that a 'live' user is indeed present at the point of data acquisition.

Multimodal systems use a combination of two or more modalities to overcome the limitations that arise when using traditional approaches and single biometric trait to recognize individuals. Multimodal biometrics system involves various levels of fusion, namely, sensor level, feature level, matching score level, decision level and rank level (Kisku *et al.*, 2009). Identity management system has a challenging task in providing authorized user with secure and easy access to information and services across a wide variety of networked system. Nguyen *et al*. (2018) also suggested that future work should concentrate on how to reduce the computational complexity of the deep learning applications because it is still an open problem. The major challenge of any deep learning approach is the fact that it is computationally complex in nature. To address this problem, Convolutional Neural Network–Genetic Algorithm (CNN-GA) architecture which applied GA, a known optimization technique was developed. The objectives of this work are to design a multimodal biometric framework which utilizes a CNN-GA for feature extraction and classification, and also to compare the performance of the Modified CNN (CNN-GA) with standard CNN using sensitivity, specificity, recognition accuracy, precision and recognition time.

2 Review of Related Works

Compound biometrics involves more than two evidences presented by different traits belonging to a user to form their identity as a whole and enhance authenticity efficiency. Although the system might be complex and cost more due to the requirement of new sensors and longer time required for recognition. Consequently, the performance can be significantly improved by utilizing more than one trait for a sole recognition. To the best of the authors' knowledge triple biometrics (where three biometric modalities are considered) is the highest biometrics that exists in the literature. Few reviews under this category are described.

.Yazdanpanah *et al.,* (2010) instigated a multimodal biometric identification system based on features extracted from three biometric modalities including face, ear and gait using Gabor and PCA. Fusion at matching score was performed on ORL face database, USTB ear database and CASIA gait database. The paper evaluated three different kinds of normalization techniques experimentally and two kinds of fusion methods. Z-score method of normalization combined with weighed product method of fusion gave the best recognition performance of 97.5% at 0.1% FAR. The innovative approach outperformed the unimodal systems on a variety of image databases.

Yaghoubi and Eliasi (2011a) proposed a multimodal biometrics using face, ear and palm based on feature extracted from visual cortex, achieved through transforming face and palmprint image with Gabor filter and ear images with Gaussian filter in HMAX method. However, matching score level fusion was employed. The classification was done using K-NN and SVM classifiers and the experimental results showed 96% accuracy rate on ORL face database, 94% accuracy rate on USTB Ear database and 96.6% accuracy rate on POLYU Palm database. Recognition time was not mentioned which is another major concern in multimodal system. The proposed approach for fusion of face, ear and palm illustrated in (Yaghoubi and Eliasi, 2011b) showed great performance using score-level fusion and SVM outclassed KNN.

Vivek *et al.*, (2012) Described feature extraction techniques of fusion fingerprint, iris and face using Gaussian Mixture Model (GMM). Expectation Maximization method was used to obtain the parameters needed to implement the GMM. The fingerprint was extracted using minutiae technique. For the iris, segmentation, normalization and feature encoding were accomplished. For the face PCA was used. The fingerprint, iris and face images were collected from FBI, poly and Sino biometrics database respectively and information were fused at match score-level using a density based score level fusion. GMM gave superior results. The performance of TMSD and PCA were compared using Euclidean distance scheme for recognition. The advantage of this classifier is its high speed (Samadi and Pourghassem, 2013).

From the work of (Snehlata *et al.,* 2014) face, ear and iris algorithms are tested individually and individual weight for face is found to be 92%, for ear 96% and iris 30%. The overall performance of the system has increased showing weight for face and ear to be 98.24% with FAR of 1.06 and FRR of 0.93 respectively. The overall performance of the system also increased with integration of the three.

Afolabi *et al.,* (2015) developed a bimodal face-fingerprint biometric security system that fuses matching scores obtained from single face and single fingerprint recognition modules for result management. A total of 270 facial images and 270 fingerprint images of students of JAF Comprehensive College, Ogbomoso were acquired using a digital camera and fingerprint scanner respectively. Principal Component Analysis (PCA) was employed to reduce the dimensions of the images. Three hundred and twenty images were used as training samples while the remaining 220 images were used as test samples. Features were extracted from face and fingerprint images using the Modified Gabor Filter feature extraction technique.

Matching scores were obtained for both face and fingerprint images using Euclidean Distance algorithm while Weighted Sum Rule fusion technique was used to compute the fused score, these techniques were implemented using MATLAB R2012a. The Performance evaluation of the system was carried out based on Overall Recognition Accuracy, Sensitivity and Specificity of the system. The results of evaluation showed that at the threshold value of 0.50, the overall recognition accuracy of the system was 91.95%; sensitivity was 91.25%, while the specificity was 100% but at the threshold value of 0.70, the overall recognition accuracy, sensitivity and specificity respectively yielded 100% performance. The results obtained implied that the developed system is highly effective in terms of the overall recognition accuracy, sensitivity and specificity of the system.

Al-waisy *et al.,* (2017) developed multimodal biometric system for personal identification based on deep learning approaches. Multiple instance of the same trait was used and fusion was done at the decision level. Strategies for handling over-fitting and generalization problems were discussed in their work. Nguyen *et al.,* (2018) worked on iris recognition with off the shelf CNN features. The work focused on the performance evaluation of some pre-trained CNN Models such as Alexnet, Googlenet, VGG, densenet, etc. The work left the problem of computational complexity as an open problem. Yanan *et al.,* (2020) automatically designed CNN Architectures using genetic algorithm for image classification

3 System Architectural Framework

The system architectural framework is divided into (5) major phases, namely: Image capturing phase, Image preprocessing phase, fusion phase, feature extraction/classification phase and decision phase. Each of the phases as shown in Figure 3.1 is succinctly discussed in subsequent sections.

3.1 Image Capturing Phase

The user face, right-ear and right-iris images were captured using VistaEY2 dual face and Iris Camera. These were consequently stored in the databases created for the three identifiers, accordingly as "user templates". The total number of face images, right-ear images and right-iris images used for training and testing the multimodal system is 1,710 images of 190 individuals. Multiple instances (i.e. 3) of the images were captured for ease of training and classification. Only 171 of the 190 individual were used to train the system while the remaining 19 were added as part of the test images to eventually measure the performance of the system. The breakdown of how the total number of captured images were selected for training and testing is as follows:

Number of Trained Images = 171 x 2 Instances of each trait x the number of traits
= 171 x 2 Instances of each trait x 3 = 1026 Images

Number of Test Images = ((171 x 1 instance of each trait) + (19 x the number of traits)) x the number of traits
= (171 + 57) x 3 = 228 x 3 = 684 Images

The number of biometric traits used is 3. It should also be noted that the design, implementation and testing of the entire system were carried out on MATLAB R2017a programming platform installed on HP corei7 processor, windows 10 Home basic workstation.

3.2 Image Preprocessing Phase

Image preprocessing is a fundamental step in image processing and computer vision. In this phase, the images were first cropped and resized to 100 x 100 pixels, and thereafter enhanced using histogram equalization algorithm. This includes primitive operations to reduce noise, contrast enhancement, image smoothing and sharpening, and advanced operations such as image segmentation.

3.2.1 Histogram Equalization Algorithm

According to (Zhu and Huang, 2012) the histogram equalization algorithm has been a conventional image enhancement algorithm for its simplicity and efficiency. It adjusts the gray level of an image according to the probability distribution function of the image and enlarges the dynamic range of the gray distribution to improve visual effects of the image. Based on the probability theory, the histogram equalization algorithm realizes the gray mapping of pixels in the image by using gray operations and transforms the histogram to one that is uniform, smooth, and has clear gray levels, so that the purpose of image enhancement can be achieved.

Suppose the gray value of the pixel in the original image is r ($0 \leq r \leq 1$) and its probability density is p(r), the gray value of the pixel in the enhanced image is s ($0 \leq s \leq 1$) and its probability density is p(s), and the mapping function is s=T(r). According to the physics meaning of the histogram, it is clear that every bar on the equalized histogram is of the same height. That is

$$P_s(s)ds = P_r(r)dr \quad (1)$$

Suppose s=T(r) is a monotonically increasing function in the interval and its inverse function r $=T^{-1}(s)$ is a monotonic function also. According to (3.1), we can deduce

$$P_s\,(s) = [P_r\,(r)\,\frac{1}{ds/dr}\,]_{r=T^{-1}(s)} = P_r\,(r)\frac{1}{Pr\,(r)} = 1 \quad (2)$$

The mapping relationship of the conventional histogram equalization algorithm: In discrete conditions, the relationship between *i* (the gray value of the pixel in the original image)
and f_i (the gray value of the pixel in enhanced the image) is

$$f_i = (m-1)T(r) = (m-1)\sum_{k=0}^{i}\frac{q_k}{Q} \quad (3)$$

where, *m* is the number of gray levels presented in the original image, q_k is the number of pixels in the image with *k* th gray level, *Q* is the total number of pixels in the image. Suppose an image has *n* different gray levels, and the occurrence probability of *i* th gray level is P_i, so the entropy of the gray level may be defined as

$$e(i) = -P_i \log P_i \quad (4)$$

The entropy of the whole image is

$$E = \sum_{i=0}^{n-1} e(i) = \sum_{i=0}^{n-1} P_i \log P_i \quad (5)$$

It can be proved that *E* will achieve its maximum if and only if $P_0 = P_1 = \ldots = P_{n-1} = \frac{1}{n}$. That is to say the entropy of the whole image achieves its maximum when the histogram of the image has uniform distribution.

From (3), it is clear that the dynamic range has been enlarged after equalization. The essence of the equalization is to expand the quantization interval.

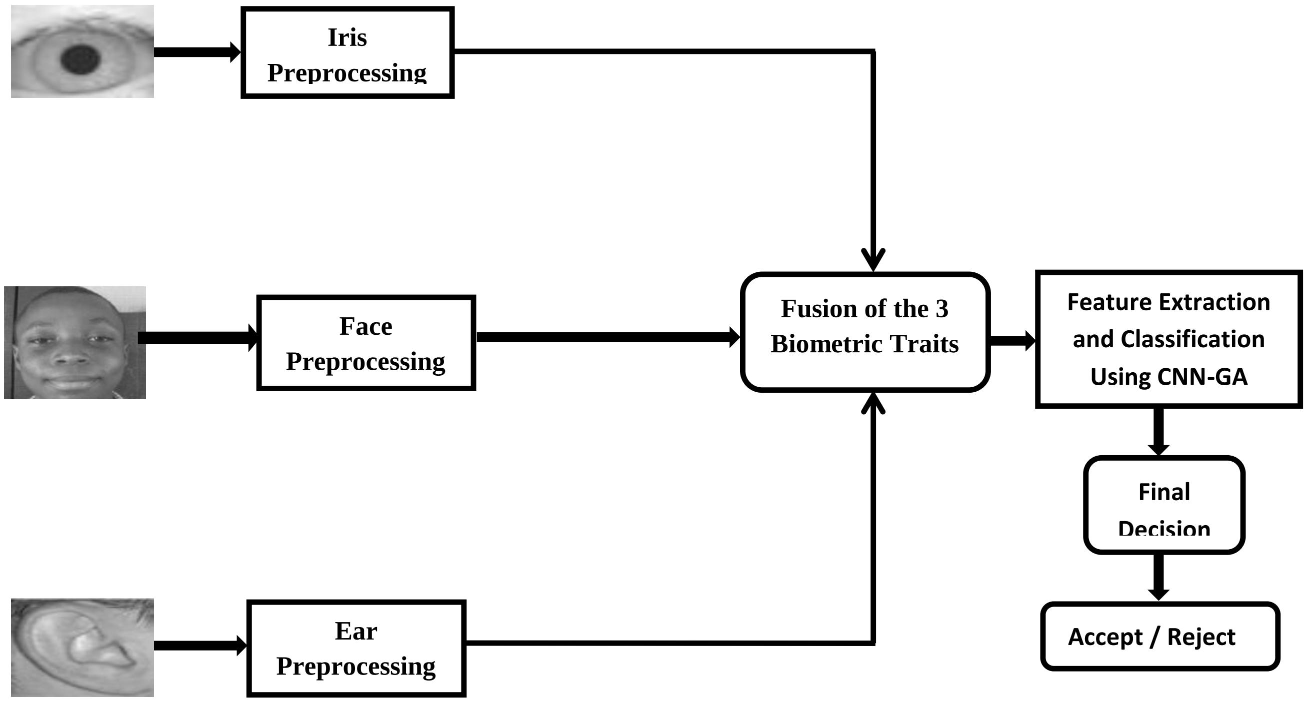


Figure 1: Architectural Framework of the Developed Multimodal Biometric Security System

3.3 Fusion Using Weighted Sum Rule Technique

Fusion of face, ear and iris modalities was carried out by applying the mathematical model given by (Pour *et al.*, 2010).

$$Weighted\ Sum = \sum_{i=1}^{n} w_i\, S_i \tag{6}$$

Where:
n is the number of preprocessed biometric modalities to be fused
S_i is the preprocessed biometric modality
w_i is the weight for each preprocessed biometric modality which can be calculated as follow:

$$Weight = \frac{EER_i}{\sum_i EER} \tag{7}$$

Where EER_i is the unimodal biometric error,
The Weighted Sum was fed into the CNN-GA model for feature extraction and classification after it might have been generated. As displayed in Figure 1, the fusion of modalities was carried out at sensor level because the amount of information available to the system gets compressed as one proceeds from the sensor module to the decision module (Ross et al., 2006).

3.4 Feature Extraction and Classification Phase Using CNN-GA

Deep learning algorithms can learn tasks directly from data, eliminating the need for manual feature election. Deep Learning is about learning multiple levels of representation and abstraction that help to make sense of data such as images, sound, and text. Deep learning performs end-to-end learning for feature extraction and classification. One of the commonly used deep learning techniques is the CNN. Once a preprocessed face, ear and iris images are obtained, and feature extraction carried out on each modality, classification shall be performed using a deep learning approach that combines Genetic Algorithm (GA) which is an optimization technique and a Convolution Neural Network (CNN).

In this work, the structure of the proposed CNN involves a combination of convolutional layers and subsampling max-pooling. The top layers in the proposed CNN are two fully connected layers for the classification task. Then, the output of the last fully connected layer is fed into the Softmax classifier, which produces a probability distribution over the N class labels.

Based on domain knowledge from the literature, there are three main aspects that have a great influence on the performance of a CNN, which need to be investigated. These include:

i. Training and Classification,
ii. Architecture / Configuration of the network and
iii. Input image size.

3.4.1 Training and Classification

The datasets were divided into training and test sets. The test sets were used to keep track of the generalization ability of the network during the learning process and storing the weights configuration that performs best on it with minimum validation error. The procedural steps followed to achieve the training and classification of modalities are as follows:

Step 1: Forward pass: output of neuron of row *k*, column *y* in the *l th* convolution layer and *k* th feature pattern in equation (3.8) among them, *f* is the number of convolution cores in a feature pattern, output of neuron of row *x* , column *y* in the *l* th sub sample layer and *k* th feature pattern in equation (3.9), the output of the *j* th neuron in *l* th hidden layer *H* in equation (3.10), among them, *s* is the number of feature patterns in sample layer, output of the *i* th neuron in *l* th layer *F* in equation (3.11).

$$O_{x,y}^{(l,k)} = \tanh(\sum_{t=0}^{f-1} \sum_{r=0}^{K_h} \sum_{c=0}^{K_w} W_{(r,c)}^{(k,t)} O_{(x+r,\ x+c)}^{(l-1,k)} + Bias^{(l,k)} \tag{8}$$

$$O_{x,y}^{(l,k)} = \tanh(W^{(k)} \sum_{r=0}^{S_k} \sum_{c=0}^{S_w} O_{(x*S_h+r,y*\,S_w+c)}^{(l-1,k)} + Bias^{(l,k)} \tag{9}$$

$$O_{(l,j)} = \tanh(\sum_{k=0}^{s-1} \sum_{x=0}^{S_h} \sum_{y=0}^{S_w} W_{(x,y)}^{(j,k)} O_{(x,\ y)}^{(l-1,k)} + Bias^{(l,j)} \tag{10}$$

$$O_{(l,i)} = \tanh(\sum_{j=0}^{H} O_{(l-1,j)} W_{(i,j)}^{l} + Bias^{(l,i)}) \tag{11}$$

Step 2: Back propagation: output deviation of the *k* th neuron in output layer *O*:

$$d(O_k^o) = y_k - t_k \tag{12}$$

Step 3: input deviation of the *k* th neuron in output layer:

$$d(I_k^o) = (y_k - t_k)\varphi(v_k) = \varphi(v_k)d(O_k^o) \tag{13}$$

Step 4: weight and bias variation of *k* th neuron in output *O*:

$$\Delta\left(W_{k,x}^o\right) = d(I_k^o)y_{k,x} \tag{14}$$

$$\Delta(Bias_k^o) = d(I_k^o) \tag{15}$$

Step 5: output bias of *k* th neuron in hidden layer *H*, *where th is the threshold*:

$$d(O_k^H) = \sum_{i=0}^{i<th} d(I_i^o)W_{i,k} \tag{16}$$

Step 6: input bias of *k* th neuron in hidden layer *H* :

$$d(I_k^H) = \varphi(v_k)d(O_k^H) \tag{17}$$

Step 7: weight and bias variation in row *x* , column *y* in the *m* th feature pattern, a former layer in front of *k* neurons in hidden layer *H*

$$\Delta\left(W_{m,x,y}^{H,k}\right) = d(I_k^H)y_{x,y}^m \tag{18}$$

$$\Delta(Bias_k^H) = d(I_k^H) \tag{19}$$

Step 8: output bias of row *x* , column *y* in *m* th feature pattern ,sub-sample layer *S*

$$d\left(O_{x,y}^{S,m}\right) = \sum_k^S d(I_{m,x,y}^H)W_{m,x,y}^{H,k} \tag{20}$$

Step 9: input bias of row *x*, column *y* in *m* th feature pattern, sub-sample layer S

$$d\left(I_{x,y}^{S,m}\right) = \varphi(v_k)d(O_{x,y}^{S,m}) \tag{21}$$

Step 10: weight and bias variation of row *x*, column *y* in *m* th feature pattern, sub-sample layer *S*

$$\Delta(W^{S,m}) = \sum_{x=0}^{fh} \sum_{y=0}^{fw} d(I_{\lfloor x/2\rfloor,\lfloor y/2\rfloor}^{S,m})O_{x,y}^{C,m} \tag{22}$$

among them, *C* represents convolution layer.

$$\Delta(Bias^{S,m}) = \sum_{x=0}^{fh} \sum_{y=0}^{fw} d(O_{x,y}^{S,m}) \tag{23}$$

Step 11: output bias of row *x*, column *y* in *k* th feature pattern, convolution layer *C*

$$d\left(O_{x,y}^{C,k}\right) = d(I_{\lfloor x/2\rfloor,\lfloor y/2\rfloor}^{S,k})W^k \tag{24}$$

Step 12: input bias of row *x*, column *y* in *k* th feature patter, convolution layer *C*

$$d\left(I_{x,y}^{C,k}\right) = \varphi(v_k)d(O_{x,y}^{C,k}) \tag{25}$$

weight variation of row *r*, column *c* in *m* th convolution core corresponding to *k* th feature pattern in *l* th layer, convolution *C*.

$$\Delta\left(W_{r,c}^{k,m}\right) = \sum_{x=0}^{fh} \sum_{y=0}^{fw} d(I_{x,y}^{C,k} O_{x+r,y+c}^{l-1,m}) \tag{26}$$

total bias variation of the convolution core

$$\Delta(Bias^{C,k}) = \sum_{x=0}^{fh} \sum_{y=0}^{fw} d(O_{x,y}^{C,k}) \tag{27}$$

Step 13: Evaluate Objective Function based on initial optimal weight features.

$$fit = \sum_{i=1}^{m} \sum_{j=1}^{n} \Delta(W_{i,j}^{m,n})((x_i) - (x_j)) \tag{28}$$

Where $\Delta(W_{i,j}^{m,n})((x_i) - (x_j)$ is the change in weight of input pixel x along the row and column

Steps 1-12 are the procedural steps involved in standard CNN. Step 13 was introduced to modify the existing standard CNN (i.e. modified CNN) for performance improvement.

3.4.2 Genetic Algorithm

Genetic Algorithm is an optimization technique that can be used to solve optimization problems. Figure 2 shows the general flowchart of a genetic algorithm:

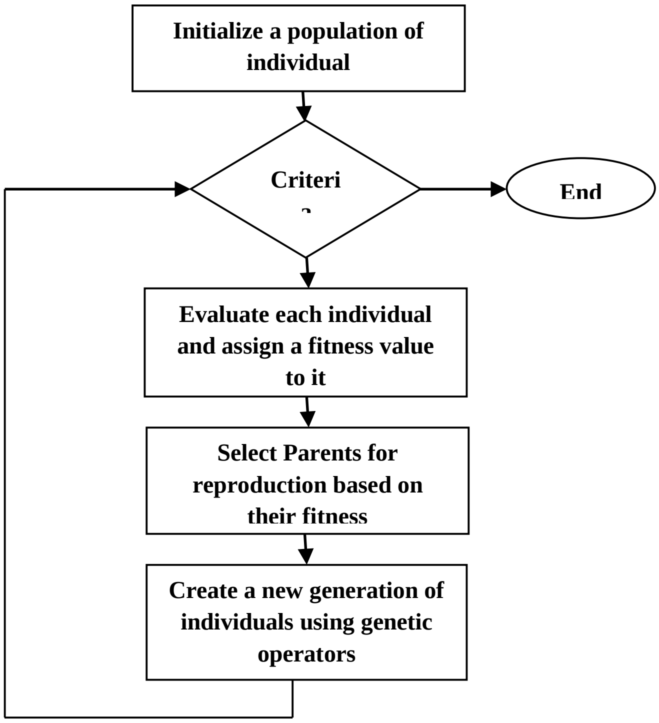


Figure 2: The general flowchart of a Genetic Algorithm

3.4.3 Network Architecture

From literatures, it appears that choosing the network architecture is still an open problem and is application dependent. The main concern in finding the best CNN architecture is the number of the layers to employ in transforming from the input image to a high-level feature representation, along with the number of convolution filters in each layer. Figure 3 shows the architecture of CNN Model that comprises of one input layer, four convolutional layers, four max-pooling layers, two fully-connected layers and one output layer.

The input layer receives the input data which in this case is the result of fusing the feature vectors of face, iris and ear modalities. The convolution layer detects the local conjunctions of features from the previous layer and maps their appearance to a feature map. There are some layers between the convolution layer and the max-pooling layer, namely: non-linearity layer and rectification layer. The non-linearity Layer which consists of an activation function takes the feature map generated by the convolutional layer and creates the activation map as its output. The rectification layer performs element-wise absolute value operation on the input volume, generally the activation volume. The max-pooling or down-sampling layer was responsible for reducing the spatial size of the activation maps while the fully-connected layers mapped the activation volume from the combination of previous different layers into a class probability distribution. It is the output layer that gave the classification result. As shown in Figure 3, the optimization of CNN using GA was done at the fully connected layers. It was at the fully connected layers that the modification CNN was introduced in this work.

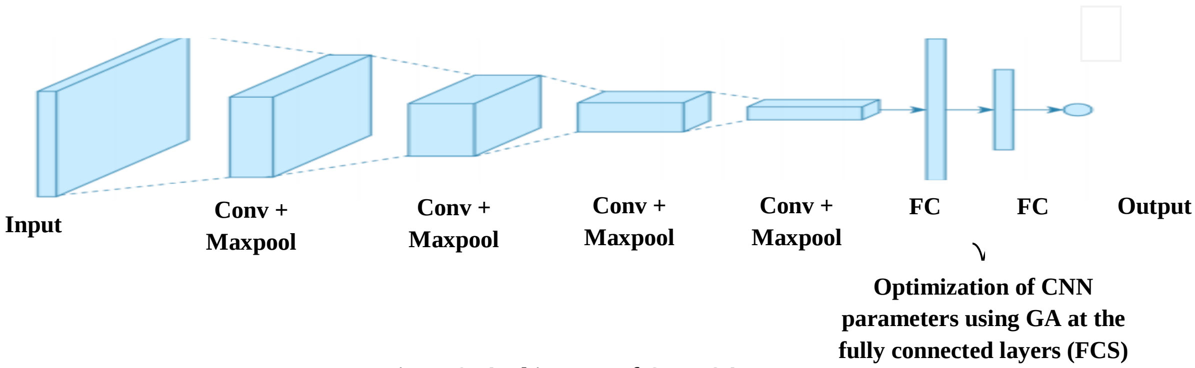


Figure 3: Architecture of CNN-GA

3.4.4 Input Image Size

The input image size is one of the hyper-parameters in the CNN that has a significant influence in the speed and the accuracy of the neural network. In this work, all captured images are resized to 100 x 100. In order to control the spatial size of the input and output volumes, a zero padding (of 1 pixel) was applied only to the input layer.

3.5 The Graphical User Interface of the Developed Multimodal System and Some Samples of the Captured Biometric Traits

A graphical user interface was designed for the developed multimodal security system for ease of experimentation. Figures 4 and 5 show the graphical user interfaces for the developed system during feature extraction and classification using standard CNN and Modified CNN respectively. The graphical user interfaces have different segments such as the training, testing, classification segments as well as other segments that display the images and display the results. Figure 6 shows a sample of original face image, cropped face image and an enhanced histogram face image while Figure 7 shows a sample of original right ear, cropped right ear image and an enhanced histogram right ear image. Figure 8 on the other hand, show a sample of original right iris image, segmented right iris image, noised right iris image and normalized right iris image.

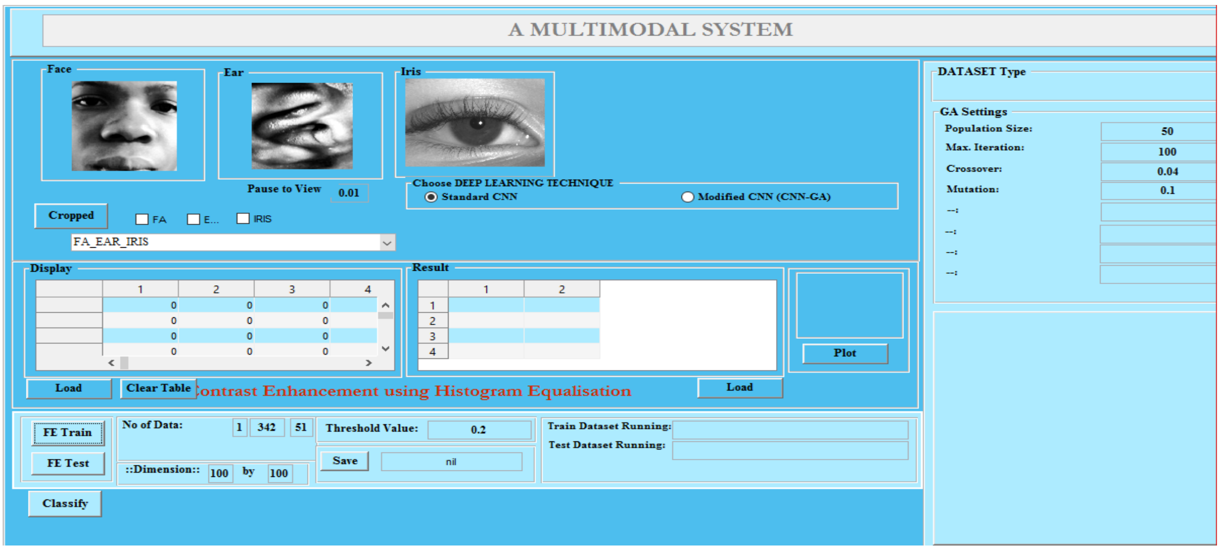


Figure 4: GUI of the Developed Adaptive Multimodal System during Feature Extraction and Classification Using Standard CNN

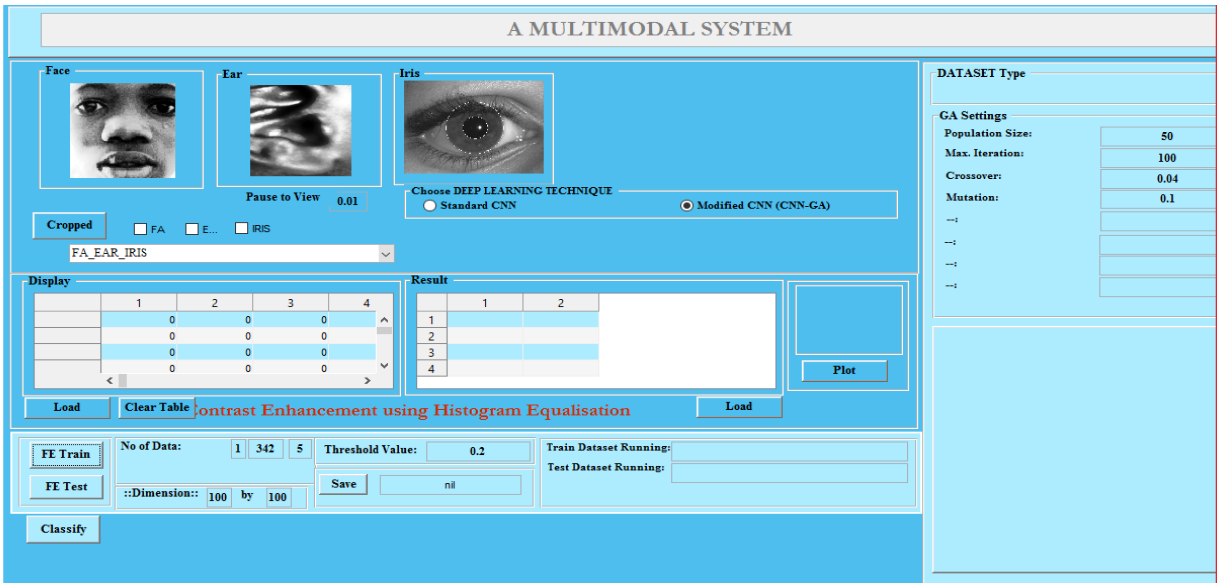
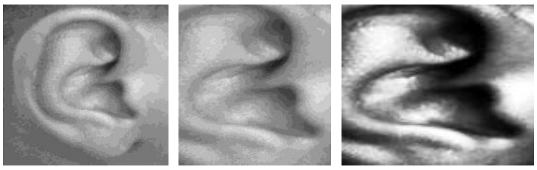


Figure 5: GUI of the Developed Adaptive Multimodal System during Feature Extraction and Classification Using Modified CNN (CNN-GA)

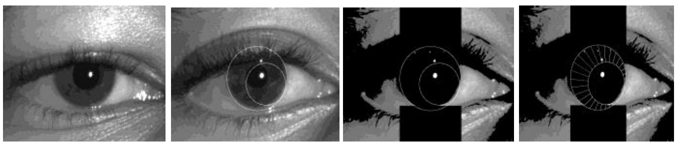
(a) (b) (c)

Figure 6: Samples of Face Images Used: (a) Original Face Image (b) Cropped Face Image (c) Enhanced Histogram Face Image



(a) (b) (c)

Figure 7: Samples of Ear Images Used: (a) Original Right-Ear Image (b) Cropped Right-Ear Image (c) Enhanced Histogram Right-Ear Image



(a) (b) (c) (d)

Figure 8: Samples of Iris Images Used: (a) Original Right-Iris Image (d) Segmented Right-Iris Image (c) Noised Right-Iris Image (d) Normalised Right-Iris Image

3.6 Performance Evaluation Metrics

The following parameters were used to measure and evaluate the overall performance of the developed system:
True Positive (TP): correctly identified images.
False Positive (FP): incorrectly identified images.
True Negative (TN): correctly rejected or unidentified images.

False Negative (FN): incorrectly rejected or unidentified images.
Sensitivity (TPR): Ability to identify presence of images in the database,

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \quad (29)$$

Specificity (TNR): Ability to identify absence of images in the database,

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100\% \quad (30)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad (31)$$

$$\text{Recognition Accuracy} = \frac{\text{TP} + \text{TN}}{\text{Total number of images}} \times 100\% \quad (32)$$

Recognition Time: It is the measure of the time-taken for the training and classification of Images.

4 EXPERIMENTAL RESULTS

Several experimental tests were conducted to validate the performance of the developed multimodal system under varying conditions. Arbitrary constants called threshold values from 0 – 1 were used to moderate the results obtained during running and testing. The general observation is that all the performance evaluation metrics gave the same results for the following ranges of threshold values respectively: 0 – 0.20, 0.21 – 0.35, 0.36 – 0.50 and 0.51 – 1. For instance, for threshold values between 0.36 – 0.50, the System accuracy, recognition time, or sensitivity, specificity, and precision of the system remains the same. This applies both to when the System was tested using CNN and CNN-GA. To aid the clarity of the results just a single threshold value is selected for each of the given ranges, namely 0.2, 0.35, 0.50 and 0.76.

4.1 Standard CNN Versus Modified CNN (CNN-GA)

BIOMETRIC TRAITS	CLASSIFIERS	TP	FN	FP	TN	Sensitivity (%)	Specificity (%)	Precision (%)	Recognition Accuracy (%)	Recognition Time (Secs)	Threshold Values
Face-Ear-Iris	CNN	167	4	11	46	97.66	80.70	93.82	93.42	560.70	**0.20**
Face-Ear-Iris	CNN-GA	170	1	9	48	99.42	84.21	94.97	95.61	418.27	
Face-Ear-Iris	CNN	166	5	9	48	97.08	84.21	94.86	93.86	561.42	**0.35**
Face-Ear-Iris	CNN-GA	169	2	6	51	98.83	89.47	96.57	96.49	426.35	
Face-Ear-Iris	CNN	165	6	7	50	96.49	87.72	95.93	94.30	562.14	**0.50**
Face-Ear-Iris	CNN-GA	168	3	3	54	98.25	94.74	98.25	97.37	436.18	
Face-Ear-Iris	CNN	164	7	4	53	95.91	92.98	97.62	95.18	565.02	**0.76**
Face-Ear-Iris	CNN-GA	167	4	1	56	97.66	98.25	99.40	97.81	455.54	

The results realized from the developed system when CNN and CNN-GA techniques were employed are as shown in Figure 4.1 at varying threshold values.

Table 1: The Performance of the Developed Multimodal System Using Standard CNN Classifier and Modified CNN Classifier in terms of recognition rates and time

4.1.1 Standard CNN Classifier Results

At the threshold value of 0.20, the True Positive records 167, False Negative records 4, False Positive records 11, True Negative records 46, Sensitivity records 97.66%, Specificity records 80.70%, Precision records 93.82%, Recognition Accuracy records 93.42% and Recognition Time records 560.70 Seconds. At the threshold value of 0.35, the True Positive records 166, False Negative records 5, False Positive records 9, True Negative records 48, Sensitivity records 97.08%, Specificity records 84.21%, Precision records 94.86%, Recognition Accuracy records 93.86% and Recognition Time records 561.42 Seconds.

At the threshold value of 0.50, the True Positive records 165, False Negative records 6, False Positive records 7, True Negative records 50, Sensitivity records 96.49%, Specificity records 87.72%, Precision records 95.93%, Recognition Accuracy records 94.30% and Recognition Time records 562.14 Seconds. At the threshold value of 0.76, the True Positive records 164, False Negative records 7, False Positive records 4, True Negative records 53, Sensitivity records 95.91%, Specificity records 92.98%, Precision records 97.62%, Recognition Accuracy records 95.18% and the Recognition Time records 565.02 Seconds.

4.1.2 Modified CNN (CNN-GA) Classifier Results

At the threshold value of 0.20, the True Positive records 170, False Negative records 1, False Positive records 9, True Negative records 48, Sensitivity records 99.42%, Specificity records 84.21%, Precision records 94.97%, Recognition Accuracy records 95.61% and Recognition Time records 418.27 Seconds. At the threshold value of 0.35, the True Positive records 169, False Negative records 2, False Positive records 6, True Negative records 51, Sensitivity records 98.83%, Specificity records 89.47%, Precision records 96.57%, Recognition Accuracy records 96.49% and Recognition Time records 426.35 Seconds. At the threshold value of 0.50, the True Positive records 168, False Negative records 3, False Positive records 3, True Negative records 54, Sensitivity records 98.25%, Specificity records 94.74%, Precision records 98.25%, Recognition Accuracy records 97.37% and Recognition Time records 436.18 Seconds.

At the threshold value of 0.76, the True Positive records 167, False Negative records 4, False Positive records 1, True Negative records 56, Sensitivity records 97.66%, Specificity records 98.25%, Precision records 99.40%, Recognition Accuracy records 97.81% and the Recognition Time records 455.54 Seconds. It was deduced from Table 4.1 that the Modified CNN yields better results than standard CNN in terms of Sensitivity, Specificity, Precision, Recognition Accuracy and Recognition Time at different threshold values. Figure 4.1 clearly shows the ROC curves of the two classifiers. The curve shows that CNN-GA yields a more excellent result than standard CNN in terms of sensitivity and specificity. For graphical illustration of the results, Figure 4.2 and Figure 4.3 shows the performance of the developed multimodal system when CNN and CNN-GA were respectively used for feature extraction and classification, the graph shows that CNN-GA thrived better in performance in terms of recognition accuracy and recognition time than CNN.

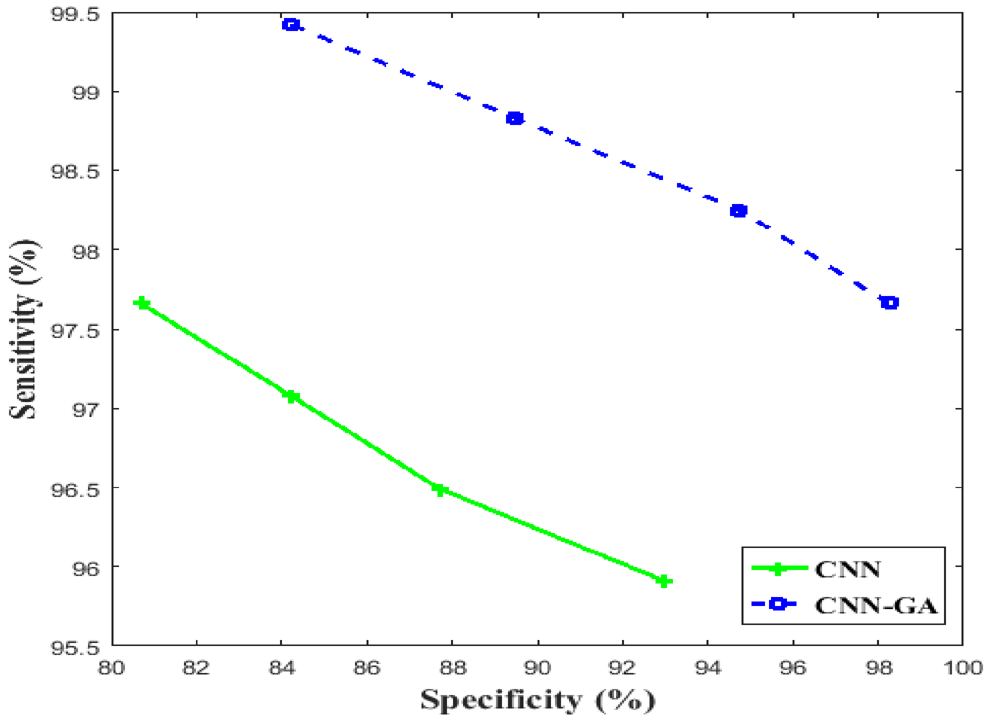


Figure 9: ROC Graph for CNN and CNN-GA Classifiers

4.2 Statistical Analysis of Some Obtained Results

Inferential statistical analysis using paired sampled t-test was done to analyze the result obtained for accuracy, recognition time and precision between CNN-GA and CNN technique of tri-modal and bi-modal biometric, and bi-modal and unimodal biometric. The test was performed to determine the level of significance in the performance of the techniques. The result obtained from the analysis using SPSS is shown in Table 2**.**

The paired sampled t-test was performed on null hypothesis (H_0) that there is no significant difference between CNN-GA and CNN technique as against the alternative hypothesis (H_1) where there is a significant difference between CNN-GA and CNN technique, at 5% level of significance. The hypothesis is defined as:

H_0: There is no significant difference between CNN-GA and CNN technique.

H_1: There is significant difference between CNN-GA and CNN technique.

From Table 2 the *p-value* for accuracy, recognition time, FPR and stability are 0.000, 0.000, 0.005 and 0.005 respectively. The *p-value* depict statistical significance at $P < 0.05$. Test of significance of the accuracy, recognition time and precision evaluated at 95% confidence level shows that there was significant difference between the CNN-GA and CNN technique. Hence the alternative hypothesis is accepted. The t-test result validates the fact that CNN-GA outperformed the CNN technique in terms of accuracy, recognition time and precision.

Table 2: Test of significance on Recognition Accuracy, time and Precision of the developed technique

Parameter	Test	T	Degree of Freedom *(df)*	p-value	Comment
Accuracy	CNN-GA vs CNN	14.641	3	0.000	Significant
Recognition Time	CNN-GA vs CNN	-18.058	3	0.000	Significant
Precision	CNN-GA vs CNN	7.272	3	0.005	Significant

Consequently, the validation of the performance of the CNN-GA technique with respect to CNN technique justifies the fact that the application of CNN along with GA technique in tri-modal biometric system based on iris, face and ear biometric gave significant improvement.

5 CONCLUSION

This work has been able to establish the fact that no single biometric information is sufficient enough to authenticate humans. The biometric information that were engaged in this work are face images, right ear images and the right iris images. All these were used because they are all passive biometrics and do not require active or full participation of individuals to be probed. Combining multiple sources of biometric information has proven to provide more reliability, accuracy and precision in as established in this work. Three biometric traits of each person were captured and integrated to enhance the performance of the multimodal biometric system that has been developed.

This work has shown that the modified CNN used for feature extraction and classification of images in the developed multimodal biometric security system outperformed the standard CNN in terms of sensitivity, specificity, precision, recognition accuracy and recognition time. Though both standard CNN and modified CNN Classifiers are deep learning approaches, the standard CNN is computationally intensive in nature. This challenge has been addressed in this work by reducing the complexities of standard CNN in terms of recognition rates and time.

5.1 RECOMMENDATIONS

Base on the results obtained from the series of experiments that have been carried out, this kind of security system is most recommended for financial institutions, educational institutions, government agencies, health sector, crime detection agencies, etc. Subsequent studies can focus on the use of more than three biometric traits, for instance four or five traits. In this work, it was only the right ear and the right iris of individuals that were adopted for human authentication but future work can concentrate on engaging both the left ear and the left iris, making the biometric system pentamodal.

Another recommendation is that instead of focusing on the face region as seen in this work, focus may be shifted to the hand region that comprises the fingerprint, palm print and finger-knuckles to further test the performance of the multimodal system. Future work can also consider integrating another optimization technique apart from genetic algorithm with CNN.

6. ACKNOWLEDGEMENTS

I appreciate the efforts of my supervisor Prof. A. O. Afolabi for the patience he exercised in reading and correcting this work. The effort of my co-supervisor Dr. A.S. Falohun in adding beauty to this work is also highly acknowledged. The Lord Almighty will increase you in all areas of life.

7. REFERENCES

Afolabi, A.O., Atanda, O.G., and Awodoye O.A. (2015). Performance Evaluation of a Modified Bimodal Biometric Security System for Student Result Management. *Proceedings of 4th International Conference on Mobile e-services, the Senate Chamber, LAUTECH, Ogbomoso, Oyo State,* pp. 68-76.

Al-Waisy, A. S., Qahwaji, R., Ipson, S., Al-Fahdawi S., and Nagem, T. A. M. (2017). A multimodal biometric system for personal identification based on deep learning approaches. *Seventh International Conference on Emerging Security Technologies (EST),* pp.163-168.

Kisku, D.R., Singh, J. K., and Tistarelli, M. (2009). "Multisensor biometric evidence fusion for person authentication using wavelet decomposition and monotonic decreasing graph. *Proceedings of 7th International Conference on advances in Pattern Recognition (ICAPR), Kolkata, India*, pp. 205-208.

Modi, S., (2011). Biometrics in Identity Management: Concepts to Applications. Artech House Information Security and Privacy. Artech House.

Nguyen, k., Fookes, C., Ross, A., and Sridharan, S. (2018). Iris Recognition with Off-the-Shelf CNN Features: A Deep Learning Perspective. Multidiscilpinary Rapid View Open Access Journal, IEEE Access 6, pp. 18848-18858.

Pour, A., Faez, K., and Amirfattahi, R. (2010): Multimodal biometric system using face, ear and gait biometrics. *10th International Conference on Information Science, Signal Processing and their Applications*.

Rahal, S. (2006). "Authentication Fingerprint System", *First National Information Technology Symposium (NITS), bridging the digital divide: challenges and solution, College of Computer & Information Sciences, King Suad University* 1, pp. 1026-1030.

Ross, A., Nandakumar, K., and Jain, K. (2006). Handbook of Multibiometrics. *Springer, New York, USA* 6, pp. 1-172.

Samadi, A., and Pourghassem, H. (2013). Children detection algorithm based on statistical models and LDA in human face images. *Proceedings of the International Conference on Communication Systems and Network Technologies, IEEE Xplore Press, Gwalior,* pp. 206-209.

Snehlata, B., Zadgaonkar, A.S., and Sinha, G.R. (2014). Multimodal Biometrics Using Face, Ear and Iris Modalities. *International Journal of Computer Applications (0975 – 8887) Recent Advances in Information Technology*, pp. 9-15.

Vivek, S.A., Aravinth, J., and Valarmathy, S. (2012). Feature extraction for multimodal biometric and study of fusion using Gaussian mixture model. *Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering, IEEE Xplore Press, Salem, Tamilnadu*, pp. 387-392.

Yaghoubi, Z., and Eliasi, M. (2011). Robust biometric authentication based on feature extracted from visual ventral stream. *Proceedings of the IEEE International Conference on Computer Applications and Industrial Electronics, IEEE Xplore Press, Penang,* pp. 448-452.

Yanan, S., Bing, X., Mengjie, Z., Gary, G., and Jiancheng, L. (2020). Automatically Designing CNN Architectures Using Genetic Algorithm for Image Classification. *IEEE Transactions on Cybernetics,* pp. 1-14.

Yanxia, W., and Qiuqi, R. (2006). Kernel fisher discriminant analysis for palmprint recognition. *Proceedings of the 18th International Conference on Pattern Recognition, IEEE Xplore Press, Hong Kong*, pp. 457-460.

Yazdanpanah, A.P., Faez, K. and Amirfattahi, R. (2010). Multimodal biometric system using face, ear and gait biometrics. *Proceedings of the Information Sciences Signal Processing and their Applications, IEEE Xplore Press, Kuala Lumpur*, pp. 251-254.

Zhu, Y., and Huang, C. (2012). An Adaptive Histogram Equalization Algorithm on the Image Gray Level Mapping . *International Conference on Solid State Devices and Materials Science,* Physics Procedia 25, pp. 601 – 608.