Received: 23 March 2018 | Revised 15 September 2023 | Accepted: 27 October 2023

Research Article
Journal of Research and Review in Science
108-123, Volume 10, December 2023
DOI:10.36108/jrrslasu/4202.11.0120

ORIGINAL RESEARCH

JOURNAL OF RESEARCH AND REVIEW IN SCIENCE

A Modified Convolution Neural Network Genetic Algorithm Multimodal Biometric Crime Control System

Ayodele O. Oloyede [1]., Ovansa A. Ateiza. [2], Vivian O. Nwaocha. [3], Taofik T. Ajagbe. [4], & Abiodun I. Aremu [5].

[1,4]Department of Computer Science, Faculty of Science, Lagos State University, Nigeria

[2,3]Department of Computer Sciences, Faculty of Sciences, National Open University of Nigeria, (NOUN) Nigeria

[5]Department of Computer Science, Lagos State University of Technology, Ikorodu, Lagos, Nigeria

Correspondence
Ayodele Oloyede, Department of Computer Science, Faculty of Science, Lagos State University, Nigeria.
Email: ayodele.oloyede@lasu.edu.ng

Funding information
Grant sponsor and grant number: NUC Research Grant (nuc.edu.ng)
*(*NUC1311/1)

Abstract:

Introduction: Globally, the rate of crime has dramatically climbed in recent years. However, a variety of the research projects are being carried out in the fields of artificial intelligence and neural networks which has necessitated unprecedented finetuning, hyperparameter optimization and large datasets.

Aim: The aim of this work is to develop a hybridized Convolution Neural Network-Genetic Algorithm (CNN-GA) model for Multimodal Biometric Crime Control System

Materials and Methods: Facial images and Thumbprint patterns used for the developed system were acquired from publicly available FG-Net and SOCOFing respectively. Procedurally, CNN and GA were used to extract facial and thumbprint features. The extracted features were fused into a single feature set using sum rule strategy and support vector machine (SVM) as classifier. The developed CNN-GA was evaluated using computational time (CT) and recognition accuracy (RA).

Results: In all, 342 images were trained and 228 images were used for testing in each of fingerprint and facial images. The result of CNN-GA on fused face and fingerprint at optimum threshold yielded RA and CT of 97.81% and 455.54s, respectively, while the corresponding values of CNN were 95.61%, and 565.02s, respectively. Also, the corresponding values of GA were 96.49% and 560.28s, respectively.

Conclusion: The developed Convolution Neural Network-Genetic Algorithm technique serves as improvement over CNN and GA in terms of recognition accuracy and computational time. This technique could be integrated into emerging crime control systems towards their improved performance.

To Keywords: Artificial Intelligence, Multimodal-Biometric, Neural-Network, Genetic-Algorithm, Facial-Images, Thumbprint-Pattern.

All co-authors agreed to have their names listed as authors.

1. INTRODUCTION

Biometrics according to [1] denotes the automatic detection of individuals with respect to their behavioral and physiological characteristics. These characteristics remain exclusive to every person and remain unaltered throughout human lifetime. One technology fast becoming the basis of an extensive array of extremely secured personal verification and identification solutions is the biometric technology. This technology, like a front end to a system, requires correct identification before it can be used or made accessible. Using biometrics aimed at personal authentication is turning out to possess a higher accuracy than traditional methods (which includes the Personal Identification Numbers – PINs or passwords use) and is less inconvenient (nothing to remember or carry). Biometrics etymology is gotten from two Greek words “bios”, this denotes life and “metron” this denotes “to measure”, thus biometrics means life measurement.

It is not only about security in biometrics, but it is likewise about user convenience. The importance of biometrics could be seen in an extensive range of military, commercial long with other related applications. Each and every aspect of our daily lives and the economy is set to be infiltrated by Biometrics. Biometric technologies, which involve mathematical analysis of a unique trait such as the retina, face, ear, iris, or fingerprint, have been widely embraced around the world [2].

Numerous forensic applications, civilian and commercial have now been implemented biometric systems as a technique for identity authentication. To determine or confirm a person's identity biometric systems are heavily reliant on evidence such as voice, face, iris, hand vein, signature, retina, fingerprints, facial thermogram, hand geometry, and so on [3]. In real-world applications, a good number of biometric systems utilized are unimodal, i.e., for verification they depend on a singular data piece for source evidence (e.g., face or single fingerprint). Multi-biometric systems can achieve the stringent performance constraints emphasized via several applications.

Multimodal biometrics system comprises numerous fusion levels, which are, sensor level, feature level, matching score level, rank level and decision level [4]. In the identity management system, a difficult process is giving an authorized user the privilege of easy and secure log on to services and information amongst an extensive variety of networked system. Several problems arise due to the disparity in numerous parameters which includes poor illumination, lighting, scale and some other environmental parameters [5]. The primary reason for biometric systems creation is basically performing binary decisions such as rejecting impostors and accepting the authorized personnel. There are two error types which are majorly bumped into in all biometric systems namely: False Rejection (FR) flaws prevent authorized persons from entering, while False Acceptance (FA) flaws enable the impersonator through.

There are many techniques that can be utilized for image classification in biometric systems. One of such techniques is the deep learning technique that has been adopted in this work. Convolutional Neural Networks (CNNs), being the leading technique of deep learning have revealed extraordinary superiority in several real-world applications over most machine learning methods [6].

2. RELATED WORKS

[7] used the t-norm technique to explore a multimodal biometric system focusing on hand features, including hand palmprint, hand veins, and geometry at score-level fusion. The testing findings demonstrated that the score-level technique using the t-norm achieved fairly decent performance and did not require any iteration. The suggested fusion employing Hama Cher t-norm at a FAR of 0% produced a Genuine Acceptance Rate (GAR) of 99.9%. As a result, there was a significant improvement in individual biometrics.

Using PCA and Gabor, [8] developed a multimodal biometric identification system based on extracted features from three biometric modalities: gait, face, and ear. On the CASIA gait database, ORL face database, and USTB ear database, fusion at matching score was done. The experimental evaluation of two fusion types methods and three different normalizing technique types was done in this paper. The Z-score normalization method combined with the weighed product method of fusion yielded the greatest recognition performance of 97.5 percent at 0.1 percent FAR. The new strategy outperformed unimodal algorithms on a range of image databases.

[9] demonstrated that when it came to face recognition, the General Regression Neural Network (GRNN) performed best, followed by Discriminant Analysis (DA), PCA, and Backpropagation Neural Network (BPNN) in that order, with DA proving to be the best technique in terms of recognition time. Among the common texture-centered feature extraction methods (Log Gabor, LPQ, LBP), LPQ for palmprint showed notable improvement, while among the common appearance-centered algorithms (LPP, LDA, PCA, ICA1) for face, LDA was shown to be the best [10]. Similar efforts were made in 2011, when the same authors used LPQ, PCA, and ICA1 for their new hybrid technique, with performance quantified in terms of EER. Furthermore, at sensor level fusion when wavelet decomposition scheme was applied, the performance was not acceptable, unimodal counterparts even had better performance.

[11] introduced a novel hybrid technique to multi-biometrics of fusion face and iris, utilizing three algorithms: Local Phase Quantization (LPQ), Independent Component Analysis1 (ICA1) and Principal Component Analysis (PCA) centered on diverse fusion level. A comparison of the suggested method with multimodal and multi-algorithmic approaches was also examined. The results of the trials were based on the FVC-2006 fingerprint database, ORL-face, and CASIA-iris, and took into account all levels of fusion except sensor level. When compared to the suggested hybrid method, EER was utilized to compare performance, and the other alternatives underperformed.

TMSD is used to locate the primary axes of variation across diverse facial expressions, and the many expression factors that are related to one another are reported, while PCA is used to depict the principal axes of variation throughout the face. The recognition of the latter surpassed the former. Faces with surprise expressions showed low identification rates because to the strong geometric alterations. In the proposed strategy, distinguishing facial expressions that appear identical, such as disgust and fury, surprise and terror, was ineffective [12].

Multimodal biometrics of fingerprint, hand vein, and iris were introduced by [13]. Experimental and theoretical investigations were compared. Score normalization procedures were analyzed and particularized, as well as the mathematical model of matching score was deduced. Weighting Average (WA) and Simple Average (SA) fusion algorithms were analyzed also. To verify the fusion theory, fingerprint database, TJU hand vein database and CASIA iris were used to access the biometric recognition experiments. The results showed that the experimental data were consistent with the deduced theoretical results.

[14] illustrated the proposed approach for fusion of Face, Voice and Signature, using score-level fusion it revealed a great performance then SVM outclassed KNN. FLDA and KFDA had substantially greater recognition rates than PCA, however FLDA (fisherpalms) performed worse than KFDA for the reason that it could not describe complicated nonlinear variations like stretching variations of hands, rotation and movement, nonetheless, it was a good feature selector because increasing the amount of training samples brings the identification rate of fisher palms closer to KFDA [15]

PCA was utilized for dimension reduction with canonical correlation analysis and higher performance was reported by [16], the PCA scheme and average rule were used for feature level fusion using LDA as the feature selector, and the PCA offset the average rule, resulting in enhanced performance. The work only focused on the extraction of a high-quality image in contactless recognition so that each user may simply lift their palms closer to the scanner; there were no experimental results for comparing the palm vein recognition distance variations results.

[17] proposed a human authentication technique that combines speech, signature, and face information in order to overcome the drawbacks of single biometric authentication, which has significant FRR and FAR issues. Based on adaptive Bayesian approach, it has established a framework for fusion of match scores in multi-modal biometric systems. The probability ratio-based fusion rule with GMM-based density achieves a notable recognition rate. A mixed authentication approach, as shown in the results, can provide a stable authentication rate and overcome the limitations of a single mode system. Based on the findings of the experiments, it was discovered that EER may be reduced greatly in the face, signature, and combination face-voice-signature modes.

The estimate algorithm of the finite Gaussian mixture model (GMM) Figueiredo-Jain (FJ) is used which is somewhat sensitive to initialization conditions and leads to obtaining different results if the algorithms is run in multiple times. Human identification systems that use automated biometrics measure a "signature" of the human body, compare the resulting characteristic to a database, and then make an application-dependent conclusion.

3. MATERIAL AND METHODS

In this work, recognition of a robust multimodal biometric crime control was performed on face and fingerprint images. A face was captured by a higher resolution pixel digital camera and Fingerprint images were captured via digital persona fingerprint sensor. In all, 342 images were trained and 228 images were used for testing in each of fingerprint and facial images.

Conversion to grayscale, image cropping, histogram equalization, binarization and thinning was used as pre-processing technique to remove noise and other unwanted elements from the captured image. Feature extraction of individual was achieved using Convolution Neural Network tuned with Genetic Algorithm while Sum rule was used to combine the attributes of features extracted from both fingerprint and facial of the test subjects. Finally, Support Vector Machine (SVM), a reliable classifier was used for the final classification and the work was implemented using Matrix Laboratory (MATLAB R2018a) Software. The metrics that were used to measure and evaluate the overall performance of the developed system were recognition accuracy, False Acceptance Rate (FAR), Equal Error Rate (EER) and False Rejection Rate (FRR). Figure 1 expressed Process flow of the developed Improved Adaptive Multimodal Crime Control System.

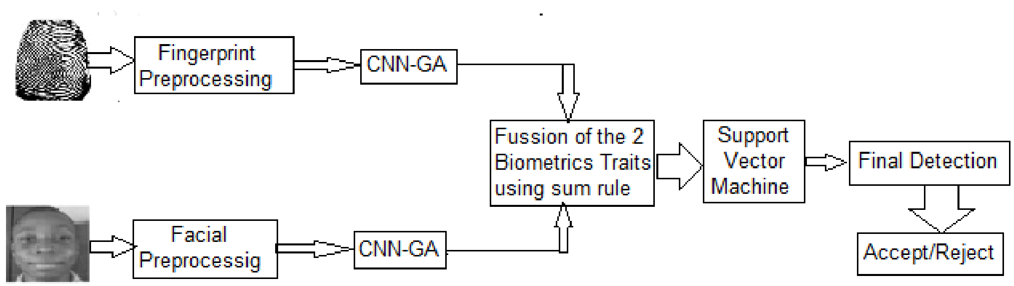

Fig. 1: Architectural design of the proposed system

3.1 Image Capturing Phase

The user’s face and fingerprint were captured using a high-definition webcam camera and fingerprint scanner. These were consequently stored in the databases created for the two identifiers, accordingly as “user templates”. Sixty percent of face images and fingerprint images were used for training and the remaining forty percent was used for testing the multimodal system.

3.2 Image Preprocessing Phase

Most times data collected is not always in a ready-to-use form. For this reason, it is expedient that the right format of data is fed into the machine learning algorithm for the problem to be solved. In view of this, the study ensured that these datasets were re-sized, converted to a grayscale, filtered, so as not to affect the image pre-processing, contrast and brightness adjustment so as to compensate the non-uniform illumination in the image. In image processing and computer vision, picture preparation is a crucial step. The photos were cropped and scaled in this phase, then enhanced using the histogram equalization process. This includes basic operations like noise reduction, contrast enhancement, image smoothing and sharpening, as well as more complex ones like image segmentation. The preprocessing techniques used for fingerprint biometrics are binarization and thinning as discussed in section 3.3.3 and 3.3.4.

3.2.1 Conversion to Grayscale

The digital camera photos were color images in three-dimensional form (3-D) that needed to be converted to grayscale (two-dimensional form (2-D) with pixel values ranging from 0 to 255. In MATLAB, each grayscale image was expressed and stored as a matrix, which was then converted to a vector image for further processing. The food vector conversion was done to help with the normalization procedure.

3.2.2 Histogram Equalization Algorithm

Because of its effectiveness and simplicity, the histogram equalization approach has remained a traditional picture enhancing algorithm, according to [18]. It enhances visual picture effects by correcting the gray level of an image corresponding to the probability distribution function of the image and then broadening the gray distribution dynamic range. The histogram equalization algorithm, which is based on probability theory, accomplishes the pixels gray mapping in the image using gray operations, then turns the histogram into one that is smooth, uniform, and has clear gray levels to meet the goal of image enhancement.

Assume the original image's gray value is r ($0 \le r \le 1$) and its probability density is p(r), the enhanced image's gray value is s ($0 \le s \le 1$) and its probability density is p(s), and the mapping function is s=T(r). Every bar on the equalized histogram is of the same height, according to the physical meaning of the histogram. That is

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

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

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

The traditional histogram equalization algorithm's mapping relationship is as follows: The relationship between *i* (the gray value of the pixel in the original image) and *fi* (the gray value of the pixel in enhanced the image) is $f_i = (m-1)T(r) = (m-1)\sum_{k=0}^{i}\frac{q_k}{Q}$ (3)

where, Q is the total number of pixels in the image, q_k is the number of pixels in the image with *k*th gray level, m is the number of gray levels presented in the original image. If an image comprises n different gray levels and the chance of occurrence of the *i*th gray level is Pi, the entropy of the gray level can be calculated as $e(i) = -Pi \log P_i$ (4)

The entropy of the whole image is $E = \sum_{i=0}^{n-1} e(i) = \sum_{i=0}^{n-1} P_i \log P_i$ (5)

It can be proved that E will achieve its maximum if and only if $P_0 = P_{12} = \ldots = P_{n-1} = \frac{1}{n}$. That is, when the histogram of the image has a uniform distribution, the entropy of the entire image reaches its maximum.

The dynamic range has been widened following equalization, as shown in (3). The goal of equalization is to make the quantization interval larger.

3.2.3 Thinning

Prior to the minutiae extraction, the final preprocessing or enhancement procedure was thinning. The morphological procedure of thinning erodes the pixels. The thinned image aids in the retrieval of minute details. The image is thinned by reducing the ridges until they are one pixel wide.

It was used to delete selected foreground pixels from binary images as a morphological process. The hit-and-miss transform is connected to the thinning procedure.

$$X-Y=X \cap \text{NOT } Y \quad (6)$$

In binary pictures, thinning is a morphological procedure that removes specified foreground pixels. It was utilized to reduce the size of the ridges to only one pixel wide by removing the unnecessary pixels. Thinning is typically used to binary images and results in the creation of another binary picture. It's the last stage before extracting the minutiae. It employed a thinning algorithm that was iterative and parallel. All pixels on the foreground region's borders that have at least one background neighbor was taken.

3.3 Fusion by Weighted Average

The pre-processed face and fingerprint features of the images were normalized using the min-max technique. The normalization of both features by the min-max rule is given by:

$$f_{face} = \frac{f'_{face} - \min(f'_{face})}{\max(f'_{face}) - \min(f'_{face})} \quad (7)$$

$$f_{finger} = \frac{f'_{finger} - \min(f'_{finger})}{\max(f'_{finger}) - \min(f'_{finger})} \quad (8)$$

where $f'_{finger}$and f'_{face} are the images obtained using pre-processed finger and face respectively, while f_{finger} and f_{face} are the normalized images. Weighted-average was used to fused the normalized images. This was achieved by using equation (9).

$$F_w = \omega f_{finger} + (1-\omega) f_{face}) \quad (9)$$

As a result, the weighted averaging method is simple to use and quick to execute. Furthermore, noise in the source photos can be suppressed via weighted averaging. It also suppresses important elements that should be kept for the fused image, resulting in a low contrast outcome. However, if the right weights are found, these problems can be solved. As a result, choosing the right value for better fused outcomes is critical.

3.4 Feature Extraction and Classification Phase Using Deep Learning Approach

Deep learning algorithms may learn tasks directly from data, removing the requirement for feature selection manually. Deep learning is the process of learning several levels of representation and abstraction to aid in the understanding of data such as images, sound, and text. For feature extraction and classification, deep learning uses end-to-end learning. The CNN is one of the most widely utilized deep learning algorithms. Once a preprocessed face, ear and iris images are obtained, and feature extraction carried out on each modality, classification was performed using a deep learning approach that combines Genetic Algorithm (GA) which is an optimization technique and a Convolution Neural Network (CNN).
The proposed CNN is built using a combination of convolutional layers and subsampling max-pooling in this study. The proposed CNN's top layers are two completely connected layers for categorization. The Softmax classifier then uses the output of the last fully connected layer to generate a probability distribution over the N class labels.

3.4.1 Training and Classification

Training and test sets were created from the datasets. The test set was utilized to track the network's generalization ability during the learning process, as well as to save the weights configuration that performs best with the least amount of validation error. The procedural steps followed to achieve the training and classification of modalities are as follows:
Step 1: Generate random population of *N*, weight space$\omega = [\omega_1, \omega_2, \dots \ \omega_n]$,Set parameter crossover probability pc , mutation probability pm
Step 2: Forward pass: output of neuron of row *k*, column *y* in the *l* th convolution layer and *k* th feature pattern in equation (11) 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 (12), the output of the *j* th neuron in *l* th hidden layer *H* in equation (13), among them, *s* is the number of feature patterns in sample layer. output of the *i*th neuron *l* th output layer *F* in equation (14).

$$O_{x,y}^{(l,k)} = \tanh\left(\sum_{t=0}^{f-1} + \sum_{r=0}^{K_h} + \sum_{c=0}^{K_w} W_{(r,c)}^{(k,t)} O_{(x+r,\ x+c)}^{(l-1,k)} + Bias^{(l,k)} \quad (10)$$

$$O_{x,y}^{(l,k)} = \tanh\left(W^{(k)} \sum_{r=0}^{S_k} + \sum_{c=0}^{S_w} O_{(x*S_h+r, y*S_w+c)}^{(l-1,k)} + Bias^{(l,k)} \quad (11)$$

$$O_{(l,j)} = \tanh\left(\sum_{k=0}^{s-1} + \sum_{x=0}^{S_h} + \sum_{y=0}^{S_w} W_{(x,y)}^{(j,k)} O_{(x,\ y)}^{(l-1,k)} + Bias^{(l,j)} \quad (12)$$

$$O_{(l,i)} = \tanh\left(\sum_{j=0}^{H} O_{(l-1,j)}^{i} W_{(i,j)}^{l} + Bias^{(l,i)}\right) \quad (13)$$

Step 3: Back propagation: output deviation of the *k* th neuron in output layer *O*:

$$d(O_k^o) = y_k - t_k \quad (14)$$

Step 4: input deviation of the *k* th neuron in output layer:

$$d(I_k^o) = (y_k - t_k)\varphi(v_k) = \varphi(v_k)d(O_k^o) \quad (15)$$

Step 5: weight and bias variation of *k* th neuron in output *O*:

$$\Delta\left(W_{k,x}^o\right) = d(I_k^o)y_{k,x} \quad (16)$$

$$\Delta(Bias_k^o) = d(I_k^o) \quad (17)$$

Step 6: output bias of *k* th neuron in hidden layer *H, where th is the threshold*:

$$d(O_k^H) = \sum_{i=0}^{i<th} d(I_i^o)W_{i,k} \quad (18)$$

Step 7: input bias of *k* th neuron in hidden layer *H* :

$$d(I_k^H) = \varphi(v_k)d(O_k^H) \quad (19)$$

Step 8: 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*

$$\Delta\left(W_{m,x,y}^{H,k}\right) = d(I_k^H)y_{x,y}^m \quad (20)$$

$$\Delta(Bias_k^H) = d(I_k^H) \quad (21)$$

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

$$d\left(O_{x,y}^{S,m}\right) = \sum_k^{170} d(I_{m,x,y}^H)W_{m,x,y}^{H,k} \quad (22)$$

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

$$d\left(I_{x,y}^{S,m}\right) = \varphi(v_k)d(O_{x,y}^{S,m}) \quad (23)$$

Step 11: weight and bias variation of row *x*, column *y* in *m* th feature pattern, sub-sample layer *S*

$$\Delta\left(W_{xy}^{S,m}\right) = \sum_{x=0}^{fh} + \sum_{y=0}^{fw} + d(I_{\lfloor x/2\rfloor,\lfloor y/2\rfloor}^{S,m})O_{x,y}^{C,m} \quad (24)$$

among them, *C* represents convolution layer.

$$\Delta\left(Bias_{xy}^{S,m}\right) = \sum_{x=0}^{fh} + \sum_{y=0}^{fw} + d(O_{x,y}^{S,m}) \quad (25)$$

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

$$d\left(O_{x,y}^{C,k}\right) = d(I_{\lfloor x/2\rfloor,\lfloor y/2\rfloor}^{S,k})W_{xy}^k \quad (26)$$

Step 13: input bias of row *x*, column *y* in *k* th feature patter, convolution layer *C*

$$d\left(I_{x,y}^{C,k}\right) = \varphi(v_k)d(O_{x,y}^{C,k}) \quad (27)$$

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

$$\Delta\left(W_{r,c}^{k,m}\right) = \sum_{x=0}^{fh} + \sum_{y=0}^{fw} + d(I_{x,y}^{C,k}O_{x+r,y+c}^{l-1,m}) \quad (28)$$

total bias variation of the convolution core

$$\Delta\left(Bias_{xy}^{C,k}\right) = \sum_{x=0}^{fh} + \sum_{y=0}^{fw} + d(O_{x,y}^{C,k}) \quad (29)$$

Step 14: Evaluate Objective Function based on initial optimal weight features.

$$fit = \sum_{i=1}^{m}\sum_{j=1}^{n}\Delta\left(W_{i,j}^{m,n}\right)((x_i) - \left(x_j\right)) \quad (30)$$

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

Step 15: Perform the following Operation

(a) Selection {as *selection pressure}*

(b) Recombination {*as the P_c used for selection of the features}*
(c) Mutation {*as the P_m used for selection of the features}*

Step 16: Generate New Selected fused features with optimal weight ω

Step 17: GOTO Step 3 until maximum iteration is reached

Step 18: Output Selected Fused Features with optimal weight ω based on best fitness value
Steps 2-14 are the procedural steps involved in standard CNN. Step 15-18 was introduced to modify and optimize the weight of CNN for performance improvement. Figure 1 showed the architectural design of an improved robust multimodal crime control system. Figure 2 described the flow diagram of training and recognition stage.

3.5 Evaluation Measures

The performance of the Convolution Neural Network tuned with Genetic Algorithm (CNN-GA) was done using recognition accuracy, False Rejection Rate (FRR), computation time, Equal Error Rate (EER) and False Acceptance Rate (FAR).

i. False acceptance Rate (FAR): The percentage of times the system admits an illegal user is known as the false acceptance rate. The rate at which imposters are mistakenly regarded as actual people is referred to as FAR. When a biometric system establishes a matching score for an impostor that meets the threshold requirements of matching, a false acceptance may result in damages. The False Match Rate (FMR) is calculated by:

$$FAR = FP/(FP+TN)X\ 100\% \qquad (31)$$

Where False Positive (FP) is the number of impostors accepted and True Negative (TN) is the number of genuine persons rejected.

ii. False Rejection Rate (FRR): The percentage of times the system rejects an authorized user is known as the false rejection rate. FRR stands for the rate at which a genuine person is accurately identified as a criminal. False Non-Match Rate is another name for FRR (FNMR).
Thus, FRR is given by

$$FRR = FN/(FN+TP) \qquad (32)$$

The number of impostors rejected is called False Negative (FN), and the number of genuine people approved is called True Positive (TP).

iii. Recognition Accuracy**:** The term "recognition accuracy" is used to describe how well a verification system performs.:
Recognition accuracy = $\big((TN+TP)TP+FN+TN+FP\big)X100\%$ (33)

iv. Equal Error Rate

3.6 Contributions to Knowledge

This research work:

i. Realized a more accurate and efficient hybridized feature extraction technique via the fusion of low-dimensional features using sum rule strategy.
ii. Vindicating CNN-GA as a better technique in bimodal than CNN in terms of accuracy.

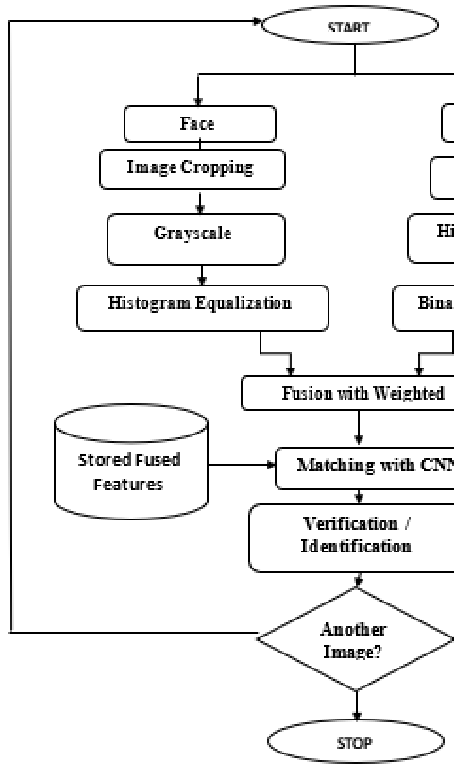

Fig. 2: Flow Diagram of Testing and Recognition Stage

4. RESULTS AND DISCUSSION

4.1 Dataset

The CNN and CNN-GA model was experimented by implementing the face and fingerprint expression recognition using 128 by 128-pixel resolution. The system was tested and evaluated using the following performance metric: sensitivity, false positive rate, recognition accuracy and computation time. All performance metrics were analyzed using a square dimension pixel resolution stated earlier at different threshold values. Three hundred and forty-two (342) face and fingerprint images were used for training which equal to 60% of the total dataset and two hundred and twenty-eight (228) face and fingerprint images which equivalent to 40% of the total dataset were using for testing. Fig 3 to 4, are the graphical user interface (GUI) showing the training phase of Fingerprint, Face and Face-Fingerprint.

Total face & fingerprint collected = 3 samples per 190 individuals
$(3\times190) = 570$
TRAINING: 2 sample $\times$ 171 individuals
$=342$ (60% of total dataset)
TESTING: 1 sample $\times$ 171 individuals + 3×19
$= 171 + 57$
$= 228$ (40% of total dataset)

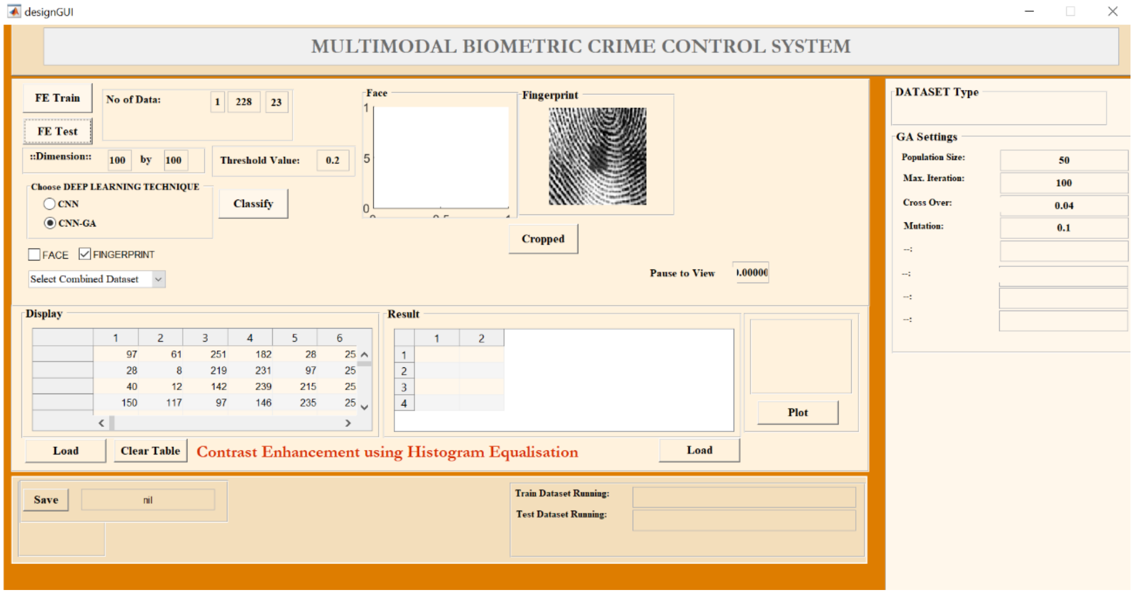

Fig. 3: Graphical User Interface (GUI) showing training phase sample of fingerprint

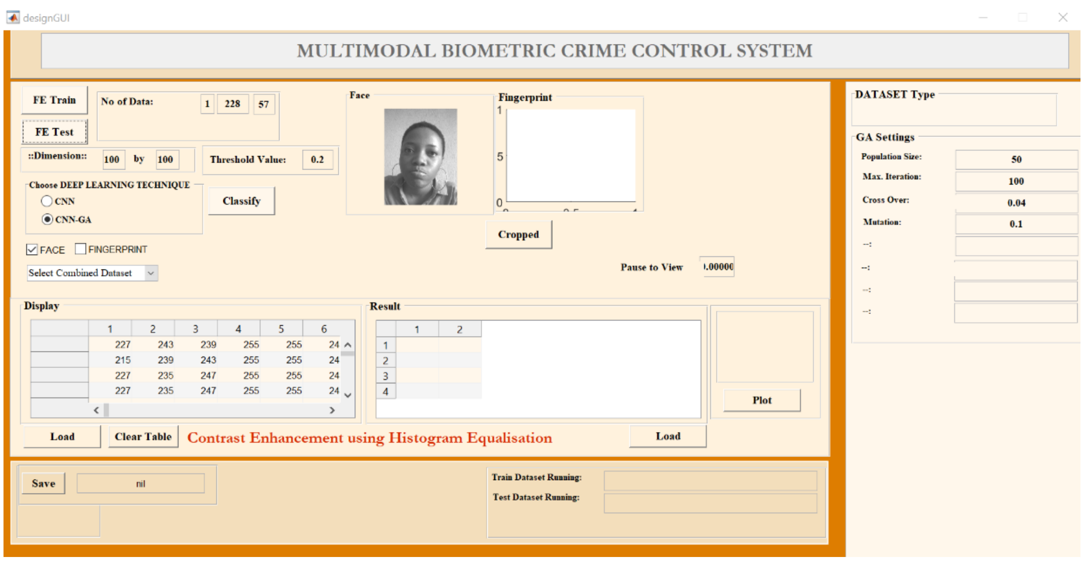

Fig. 4: Graphical User Interface (GUI) showing training phase sample of face

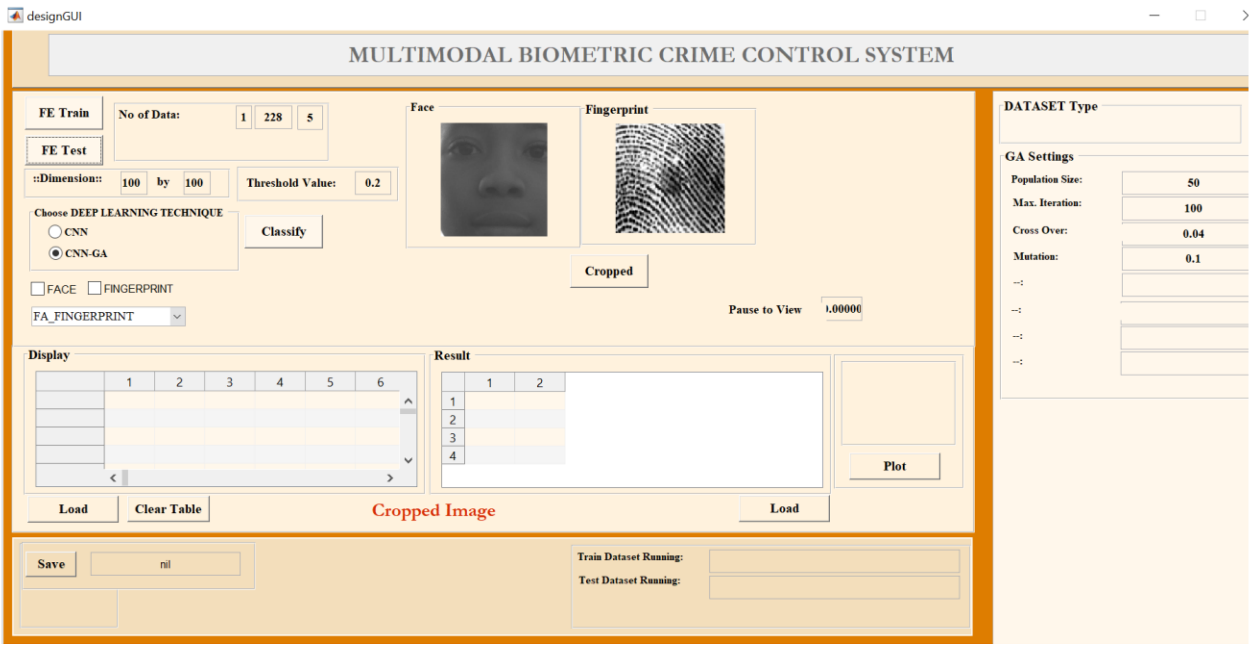

Fig. 5: Graphical User Interface (GUI) showing training phase sample of face and fingerprint

4.1.1 Hardware and Software Requirement

The algorithms were implemented using MATLAB software on windows 10 pro, 64-bit operating system (O.S), 500GB HDD, Intel® Celeron®, Central Processing Unit (CPU) N3060 @ the speed of 1.60GHz and 8GB Random Access Memory (RAM).

4.2 Evaluation Results (Training)

The dataset used contain 570 liveness fingerprint images and 570 liveness face images, 342 of the fingerprint images and 342 of the face images are used in training the model while 228 of the fingerprint images and 228 of the face images are used to test the model. The training is carried out using CNN-GA and CNN with fingerprint, face and fused fingerprint and face.

4.3 Results for Fingerprint

Table 1 described the result gotten by the Fingerprint with CNN-GA while Table 2 describes the result gotten with CNN both at threshold value of 0.2, 0.35, 0.5 and 0.76 with respect to the performance metrics. The results attainable from tables reveals that at threshold value of 0.76, the introduction of Fingerprint with CNN-GA and CNN realized a false acceptance rate of 5.26% and 7.02% respectively, false rejection rate of 5.26% and 8.16% correspondingly and an accuracy of 94.74% and 92.11% at 228.30s and 323.89s respectively. The computation time ranges between 228.03s to 231.46s and 319.62s and 323.89 seconds. Figure 6 demonstrate different recognition accuracy at different set of thresholds. The choice of some selected threshold used in this research work was based on the optimum accuracy gotten at different threshold range as shown below.

Table 1: Fingerprint with CNN-GA

TP	FN	FP	TN	FAR (%)	FRR (%)	ACC (%)	Time(sec)	Threshold
167	4	12	45	21.05	2.34	92.98	228.03	0.20
165	6	9	48	15.79	3.51	93.42	231.46	0.35
163	8	5	52	8.77	4.68	94.30	225.19	0.50
162	9	3	54	5.26	5.26	94.74	228.30	0.76

Table 2: Fingerprint with CNN

TP	FN	FP	TN	FAR (%)	FRR (%)	ACC (%)	Time(sec)	Threshold
160	11	11	46	19.30	6.43	90.35	323.85	0.20
159	12	9	48	15.79	7.02	90.79	322.48	0.35
158	13	7	50	12.28	7.60	91.23	319.62	0.50
157	14	4	53	7.02	8.19	92.11	323.89	0.76

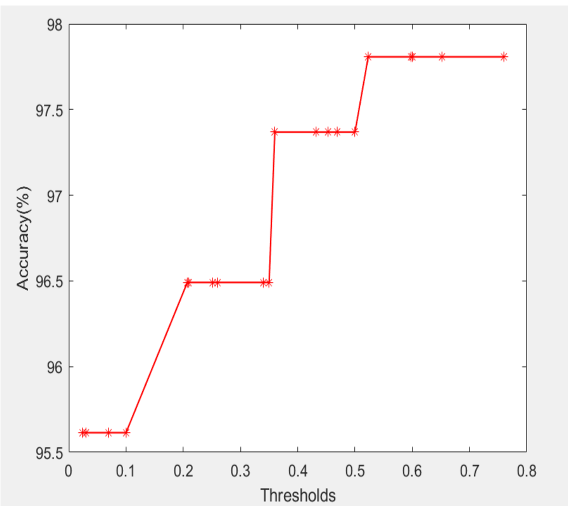

Fig. 6: Recognition Accuracy of CNN-GA against some experimental thresholds

4.4 Results for Face

Similarly, table 3 and table 4 presented the results obtained by the Face with CNN-GA and CNN correspondingly at threshold values of 0.2, 0.35, 0.5 and 0.76 with respect to the performance metrics.
The results obtainable from table 4.3 reveals that at the threshold value of 0.76, the application of Face with CNN-GA had a false acceptance rate of 3.51%, false rejection rate of 3.51% and accuracy of 96.49% at 297.01 seconds. The table 4 also shows that the computation time ranges between 298.34 to 297.01 seconds. While the results obtainable from table 4 reveals that at threshold value of 0.76, the application of Face with CNN had a false acceptance rate of 7.02%, false rejection rate of 7.60% and accuracy of 92.54% at 396.25 seconds. The table 4 also shows that the- computation time ranges between 392.35 to 399.41 seconds.

4.5 Results for fused face and fingerprint

Table 5 and 6 presented performances evaluated based on recognition accuracy, false acceptance rate and false rejection rate with respect to application of CNN-GA and CNN on fused fingerprint and face. The accuracies generated by fused fingerprint and face were analyzed at threshold values of 0.2, 0.35, 0.5 and 0.76 respectively.

Out of all threshold values considered as obtainable in Table 5 and 6. it was noticed that recognition accuracy with introduction of fused fingerprint and face at threshold value 0.76 and above was 97.81% and 95.61% higher in values than other thresholds. Hence, the fused fingerprint and face at 0.76 threshold performed better in accuracy for both tables, but had 1.75% false acceptance rate on table 5 and had 3.51% false acceptance rate on table 6 then 2.34% false rejection rate on table 5 along with a false rejection rate of 4.68% on table 6.

Table 3: Face with CNN-GA

TP	FN	FP	TN	FAR (%)	FRR (%)	ACC (%)	Time(sec)	Threshold
168	3	11	46	19.30	1.75	93.86	298.34	0.20
167	4	8	49	14.04	2.34	94.74	292.87	0.35
166	5	5	52	8.77	2.92	95.61	297.75	0.50
165	6	2	55	3.51	3.51	96.49	297.01	0.76

Table 4: Face with CNN

TP	FN	FP	TN	FAR (%)	FRR (%)	ACC (%)	Time(sec)	Threshold
162	10	11	46	19.30	5.81	90.83	399.41	0.20
160	11	9	48	15.79	6.43	91.23	393.16	0.35
159	12	7	50	12.28	7.02	91.67	392.35	0.50

Table 5: Fused Fingerprint and face with CNN-GA Result

TP	FN	FP	TN	FAR (%)	FRR (%)	ACC (%)	Time(sec)	Threshold
170	1	9	48	15.79	0.58	95.61	418.27	0.20
169	2	6	51	10.53	1.17	96.49	426.35	0.35
168	3	3	54	5.26	1.75	97.37	436.18	0.50
167	4	1	56	1.75	2.34	97.81	455.54	0.76

Table 6**:** Fused Fingerprint and face with CNN Result

TP	FN	FP	TN	FAR (%)	FRR (%)	ACC (%)	Time(sec)	Threshold
166	5	9	48	15.79	2.92	93.86	552.70	0.20
165	6	7	50	12.28	3.51	94.30	561.42	0.35
164	7	5	52	8.77	4.09	94.74	562.14	0.50
163	8	2	55	3.51	4.68	95.61	565.02	0.76

4.6 Discussion of Results

The results obtainable in Table 1 - Table 6 shows the performance of the techniques employed in this project. The results show that there is significant variation in the performance metrics with increase in threshold value and the best result is obtained at the threshold value of 0.76 across all metrics (false rejection rate, false acceptance rate and accuracy) for fused Fingerprint and Face, Fingerprint and Face respectively. Therefore, the performance of the developed technique is more dependent on the threshold value.

The fused fingerprint and face gave 97.81%, Fingerprint had 94.74% and Face 96.49% got recognition accuracies with CNN-GA respectively. While, the fused fingerprint and face gave 95.61%, Fingerprint had 92.11% and Face got 92.54% recognition accuracies with CNN respectively. It can be inferred from the results based on the performance metrics that CNN-GA applied with fingerprint and face gave the best result.

Recognition accuracies coupled with false acceptance rate generated with fused fingerprint and face with CNN-GA at 0.76 threshold values are as follows: fused fingerprint and face generated 97.81% accuracy at 1.75% FAR, Face had 96.49% accuracy at 3.51% FAR and Fingerprint got 94.74% at 5.26% FAR respectively. While with CNN at 0.76 threshold values are as follows: fused fingerprint and face generated 95.61% accuracy at 3.51% FAR, Face had 92.54% accuracy at 7.02% FAR and Fingerprint got 92.11% at 7.02% FAR respectively.

Finally, the aforementioned results were determined based on the optimum threshold value which happened to be selected because of its outstanding performance compared to other threshold values. In view of the above results, fused fingerprint and face with CNN-GA gave more accurate results due to high number of true positive as well as low number of true negative leading to high accuracy.

4.6.1 Comparison of Result between CNN and CNN-GA

Table 7 illustrate a combined result of CNN and CNN-GA at the threshold value of 0.76 with respect to all matrices at 128 by 128-pixel resolution. All result obtained in the Table 7 presume that CNN-GA model has the lowest recognition time compared with the corresponding CNN model irrespective of the threshold value.

Similarly, Recognition accuracy, sensitivity, false positive rate and specificity of CNN and CNN-GA model were compared at 128 by 128 dimensional sizes, the study discovered that CNN-GA model has better performance in accuracy, specificity and false positive rate than CNN model as enumerated in Table 5, the recognition accuracy of 97.81% with CNN-GA and 95.61% with CNN model at a threshold of 0.76 respectively.

Table 7: comparison of CNN and CNN-GA at 128 by 128-pixel resolution at 0.76 threshold value

Algorithm	TP	FN	FP	TN	FAR (%)	FRR (%)	ACC (%)	Time(sec)	Threshold
CNN-GA	167	4	1	56	1.75	2.34	97.81	455.54	0.76
CNN	163	8	2	55	3.51	4.68	95.61	565.02	0.76

5. CONCLUSION AND FUTURE WORK

5.1 Conclusion

This project evaluated the essential features of unimodal (fingerprint and Face) and multi-modal (fused fingerprint and face) on the performance of multi-modal biometrics system. Three hundred and forty-two (342) images were trained, two hundred and twenty-eight (228) images were used to test in each of fingerprint and facial images.

The experimental results obtained revealed that the fused fingerprint and face under CNN-GA gave 97.81% in terms of recognition accuracies, 1.75% false acceptance rate, 2.34% false rejection rate, and 455.54s recognition time compare with fingerprint and face modality. In view of this, an automated bi-modal system based on fused fingerprint and face (that is, both face and fingerprint), would produce a more reliable accurate and secure bi-modal system on any repository system as a result of its high accuracy. In other words, the developed CNN-GA technique has ensured good imperceptibility and classification performance, and robustness against various attacks with optimal computationally efficiency in terms of its accuracy and time. This study has contributed to knowledge by developing an improved bimodal-based access control system that was able produced a robust security system.

5.2 Future Work

With regard to the performance of the developed technique; SVM based fingerprint system can be used to enhance security challenges in an automated machine such as ATM.
It is recommended that:

i. Some evolutionary search algorithm such as Ant Colony Optimization (ACO), Evolutionary Programming (EP), GLCM (GP), Differential Evolution (DE), Artificial Immune Systems (AIS), can be introduced as feature selection techniques in other to aid recognition process.
ii. Other Artificial Neural Network techniques could be compared with SVM in other to determine its computational efficiency on fingerprint systems.
iii. A computer system with higher configurations and capability should be employed in other to handle more datasets because test-running the system with large dataset took a longer time to process.

References

[1] Oloyede, A.O., Akinola, W.A., Nwaocha Vivian. O, Longe, O. & Ugwoke F.N. (2020): An Adoption of Minutiae-Based Feature Extraction Technique for a Finger Vein Recognition System. Advances in Multidisciplinary & Scientific Research. Vol. 6, No. 1. pp 47-58. *https://www.isteams.net/aimsjournal*

[2] Abhijit, V. N., and Virani, H. G. (2016). Multimodal Biometric System Using Fingerprint, Iris and Ear. *International Journal of Technology and Science,* 9(1), pp. 40-45.

[3] 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, Ogbomosho, Oyo State*, pp. 68-76.

[4] Kisku, D., Gupta, P., and Sing, J. (2011). “Multibiometric Feature Level Fusion by Graph Clustering. *International Journal of Security and Its Applications*, 5(2), pp. 61.

[5] Jing, X., Yao, Y., Zhang, D. and Li, M. (2007). Face and palmprint pixel level fusion and Kernel DCV-RBF classifier for small sample biometrics recognition. *Pattern Recognition*, 4(11), pp. 3209–3224.

[6] Krizhevsky, A., Sulskever, I., and Hinton, G. (2012). ImageNet classification with deep convolutional neural networks. *In: Advances in neural information processing system,* pp. 1097–1105.

[7] Hanmandlu, M. J., Grover, V.K., and Vasirkala, S. (2010). Score level fusion of hand-based biometrics using t-norms. *Proceedings of the International Conference on Technologies for Homeland Security, IEEE Xplore Press, Waltham, MA*, pp. 70-76.

[8] Turki, A.A. (2016). Evaluation of Multimodal Biometrics at Different Levels of Face and Palm Print Fusion Schemes. *Asian Journal of Applied Sciences* 9, pp, 126-130.

[9] Ahmad, T., Jameel, A., and Ahmad, B. (2011). Pattern recognition using statistical and neural techniques. *Proceedings of the International Conference on Computer Networks and Information Technology, IEEE Xplore Press, Abbottabad*, pp. 87-91.

[10] Imran, M., Rao, A., and Kumar, G.H. (2013). Multimodal biometric fusion of face and palmprint at various levels. *Proceedings of the International Conference on Advances in Computing, Communications and Informatics, IEEE Xplore Press, Mysore*, pp. 1793-1798.

[11] Imran, M., Rao, A., and Kumar, G.H. (2011). A new hybrid approach for information fusion in multibiometric systems. *Proceedings of the 3rd National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics, IEEE Xplore Press, Hubli, Karnataka*, pp. 235-238.

[12] Jacey-Lynn, M., and Duncan, G. (2011). A tensor-based multivariate statistical model for 3D face and facial expression recognition. *Proceedings of the 7th International Conference on Information Technology in Asia, IEEE Xplore Press, Kuching, Sarawak*, pp. 1-8.

[13] Xiuyan, L., Changyun, M., Tiegen, L., and Chenhu, Y. (2011). “Theoretical analysis and experimental study on multimodal biometric”, Proceedings of the International Conference on Control, Automation and Systems Engineering, IEEE Xplore Press, Singapore, pp. 1-4.

[14] Soltane, M. (2015). Face, Voice and Signature Multi-modal Biometric Verification Fusion Systems. *International Journal of Engineering* 4, pp. 139-150.

[15] Tekade, A.A. (2012). Feature fusion method based on fisher discriminant analysis for face and ear for multimodal recognition. International Journal of Engineering Res. Technology.

[16] Wang, Z., Liu, C., Shi, T., and Ding, Q. (2013). Face-palm identification system on feature level fusion based on CCA. Journal of Information Hiding Multimedia Signal Processing 4, pp. 272-279.

[17] Soltane, M. (2017). Product of Likelihood Ratio Scores Fusion of Face, Speech and Signature Based FJ-GMM for Biometrics Authentication Application Systems. *Mathematics and Computer Science*. 2(5), pp. 51-65. DOI: 10.11648/j.mcs.20170205.11

[18] 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.