Face Recognition based on Convolution Neural Network and Scale Invariant Feature Transform

Jamilah ALAMRI[1], Rafika HARRABI[2], Slim BEN CHAABANE[3]
Faculty of Information Technology, Department of Information Technology
Industrial Innovation and Robotics Center
University of Tabuk, Tabuk
Kingdom Saudi Arabia

***Abstract*—Recently, Face Recognition (FR) has been received wide attention from both the research community and the cyber security industrial companies. Low accuracy of recognition is considered a main challenge when it comes to talking about employing the Artificial Intelligence (AI) for FR. In this work, the Scale Invariant Feature Transform (SIFT) and the Convolutional Neural Networks (CNN) feature extraction methods are utilized to build an AI based classifier. The CNN extracts features through both the convolutional and polling layers, while the SIFT extracts features depending on the scale space, directions, and histograms of points of interest. The features that are extracted by the CNN and the SIFT methods are used as an inputs for the KNN classifier. The experimental results with 400 test images of 40 persons, with 240 images are randomly chosen as training sets and 160 images from test sets, demonstrate in terms of accuracy, sensitivity, and error rate, that the CNN-based KNN classifier achieved better results when compared to the SIFT-based KNN classifier (accuracy = 97%, sensitivity = 93%, error rate = 3%).**

Keywords*—Face recognition; training; testing; CNN; SIFT; accuracy; classifier

I. Introduction

Background The Face Recognition (FR) research field can be seen as an intersection of three main domains, which are Artificial Intelligence (AI), Image Processing (IP), and Cybersecurity (Cs). Fig. 1 illustrates the face recognition research field in terms of domains' intersection.

For the Artificial Intelligence (AI), it is defined as the science that addresses the mechanisms of learning machines to be able to make decisions as the human's brain [1]. The Image Processing (IP) research field is defined as method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it [2]. Cybersecurity is defined as the mechanisms that are employed to protect digital data against unauthorized network users or malicious alternations [3].

Motivation (importance of domain). In the context of Smart Cities (SCs), FR-based systems play a significant role to perform tasks easily and quickly for the users. FR-based systems can save the user's time. For example, instead of opining door using keys, the FR-based system can do this mission directly once the user stands in front of the door. This saves the time of the user when he or she forgets the keys of the door [4]. Moreover, FR-based systems ensure performing the tasks at a high level of security. That is because nobody can login to sensitive locations (or data) if the system denies the matching process [5]. Furthermore, from medical point of view, FR-based systems contribute to limit the spread of Covid-19. That is because fingerprinting-based systems can be replaced by FR-based systems [6]. Actually, it is recommended not to use fingerprinting systems in both governmental and private institutions (PMC, 2020).

Statement of problem. In terms of cybersecurity, authentication security requirement means the process of identifying the identity of a user with guaranteeing that no impersonation [7]. FR contributes to provide authentication for users by processing the image of the user's face and then matching it with what was stored in a database. However, employing the Artificial Intelligence (AI) to build FR-based system is critical especially when it comes to talking about logging in to a top-secret data centers, such as the servers' room in an interior ministry, or any kind of digital information [8]. That is because any error in the FR-based system leads to open the door for attackers (unauthenticated users) [9]. This in turn means a very critical security gap in the system.

This reflects the importance of providing FR-based systems with a high accuracy. Otherwise, a big security problem will occur. Fig. 2 illustrates the problem of low accuracy of FR-based systems from cybersecurity perspective.

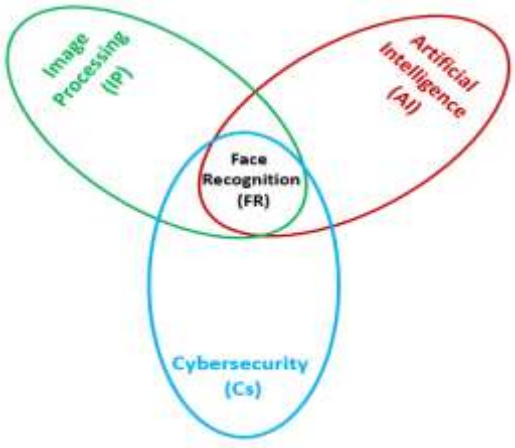

Fig. 1. Face Recognition Research Field in Terms of Domains' Intersection.

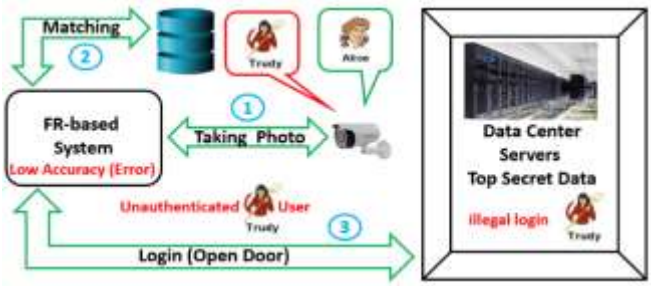

Fig. 2. The Problem of Low Accuracy of FR-based Systems.

In Fig. 2, there are two users (Alice and Trudy). Alice is authenticated user, while the Trudy is un-authenticated one. The FR-based system is linked to a data center that includes servers where a top-secret data is stored in it. Three main steps are required to legal login to the data center. First, the camera takes a photo of the face of the user. Second, the FR-based system processes the image and performs a matching process. If the information extracted from the processed image (the face of the user is recognized) matches with what is stored in the data base, the login is performed (physically, the door is opened). In the case of low accuracy, the FR-based system will have a security gap. This security gap can be exploited by the Trudy to gain an illegal login.

Research questions. There are some critical reasons of the low accuracy in FR-based systems. All of the reasons have a one root related to blurring. From the term of blurring, a main research question can be derived, which is how to ensure robustness against blurring images of the faces and high accuracy of face recognition at the same time? In details, we have the following tow research questions:

1) How to ensure high face recognition rate under the impact of noisy images of faces (such as wet faces or sweaty face).

2) How to guarantee high accuracy when dealing with different cases where the directions of the faces cause some distortion of the face.

By employing Convolution Neural Networks (CNN), we can response to the research questions. We can exploit the structure of the CNN that contains constructing the convolution and pooling layers to enhance the processing of the input images. In addition, we can support the CNN by a strong pre-processing step to ensure high resistance against noisy images.

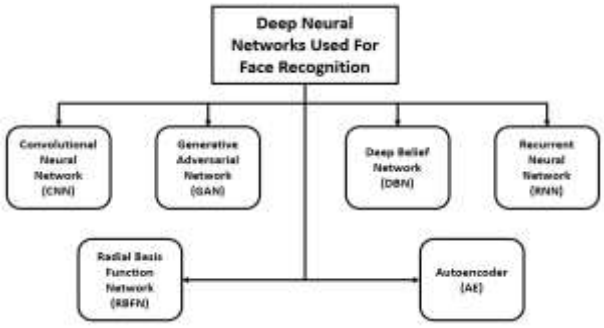

Fig. 3. Groups of Deep Learning Networks used for FR-based Systems.

Contribution. In general, the contribution of this work is as follows:

- In responding to the first research question, an efficient pre-processing step is conducted before starting the process of training the classifier. One of the aims is removing noise.
- To deal with the distortion caused by looking at different directions, this work presents the Convolutional Neural Network (CNN) and the Scale Invariant Feature Transform (SIFT) methods for feature extraction. Integration of both the CNN and the SIFT with the K-Nearest Neighbor (KNN) classifier ensure high level of accuracy under various directions of the face being classified.
- Extensive experiments are conducted to proof the effectiveness of the proposed classifiers.

Structure of paper. The rest of the work is organized so that in Section II we present the related work. Section III provides the proposed system. In Sections IV and V, the used metrics and the experiments and evaluations are conducted, respectively. Finally, the work is concluded in Section VI.

II. Related Work

In general, the Artificial Intelligence (AI) provides significant contribution to enhance the FR-based systems. Under the umbrella of the Artificial Intelligence (AI), the deep learning networks that are used for building FR-based systems can be classified into six groups, as shown in Fig. 3.

In [10] a suggested algorithm was proposed to increase the efficiency of the Elman neural algorithm in face recognition. The proposed algorithm was studied on the images of 20 students from the Department of Computer Science, Tikrit University. First step creates dataset of faces, second step convert color space to HSI and using saturation layer, image decomposition using curve let transform, feature extraction using Principle component analysis, and final step face recognition using Elman neural network. After applying proposed algorithm, the rate of face recognition 94%."

In their work [11], the authors proposes an algorithm for face detection and recognition based on convolution neural networks (CNN), which outperform the traditional techniques. In order to validate the efficiency of the proposed algorithm, a smart classroom for the student's attendance using face recognition has been proposed. The face recognition system is trained on publically available labeled faces in the wild (LFW) dataset. The system can detect approximately 35 faces and recognizes 30 out of them from the single image of 40 students. The proposed system achieved 97.9% accuracy on the testing data. Moreover, generated data by smart classrooms is computed and transmitted through an IoT-based architecture using edge computing. A comparative performance study shows that our architecture outperforms in terms of data latency and real-time response.

In [12], an efficient face recognition method using AGA and ANFIS-ABC has been proposed. At first stage, the face images gathered from the database are preprocessed. At

Second stage, an interest point which is used to improve the detection rate consequently. The parameters used in the interest point determination are optimized using the Adaptive Genetic Algorithm. Finally using ANFIS, face images are classified by using extracted features. During the training process, the parameters of ANFIS are optimized using Artificial Bee Colony Algorithm (ABC) in order to improve the accuracy. The performance of the proposed ANFIS-ABC technique is evaluated using an ORL database with 400 images of 40 individuals, YALE-B database with 165 images of 15 individuals and finally with real time video the detection rate and false alarm rate is compared with proposed and existing methods to prove the system efficiency.

In [13] the authors have presented the feature-based method for 2D face images. Speeded up robust features (SURF) and scale-invariant feature transform (SIFT) are used for feature extraction. Five public datasets, namely Yale2B, Face 94, M2VTS, ORL, and FERET, are used for experimental work. Various combinations of SIFT and SURF features with two classification techniques, namely decision tree and random forest, have experimented in this work. A maximum recognition accuracy of 99.7% has been reported by the authors with a combination of SIFT (64-components) and SURF (32-components).

In [14], introduced a method to gain the invariant illumination signs of face images based on logarithmic fractal dimension with respect to complete 8-local dimensional patterns. This method depended on performing three tasks identified by using adaptive holomorphic filter to shrink the illumination partly. Second, implement the abstracted LFD method to improve facial aspects. Third, employ the full ELDP (CELDP) that utilizes the directions and the magnitude of the edge to generate the term of illumination invariant representation. The realized results based Yale B, extended Yale B and AR achieved the database results depending on their applications. The proposed method demonstrated colossal recognition excellence by reaching the entire face recognition accuracy by 99.47% for Yale B, 99.53% for CMU-PIE, 94.55% for extended Yale B, and 86.63% for AR face databases.

The author in [15] furthermore, [14] presented krawtchouk polynomial moments technique based methodology for local descriptor. Based on edge indicator, canny edge was employed to discover the focused points to specify the zone near its scale and normalize the relation. The krawtchouk polynomial will be applied on the realized region in order to construct the descriptor. The output of the ORL, FERET based method emphasized that the results were perfect confirmed by accuracy rate of (97.86) percent.

Another technique presented in [16] named Coupled Marginal Discriminant Mappings (CMDM), which matches the images of the face with different clarity levels regardless the conditions of global data distribution and local data structure based learning map. The accuracy results obtained based on AR and FERET were realized by (94.56) and (88.5) percent respectively. Additionally, dimensionality reduced local directional pattern (DRLDP) approach proposed by [17] which showed eight-bit code assigned to (3 × 3) of every sub zone. The code describes the textural pattern of the whole block and then obtains a sole eight-bit code for each block. Experiments were performed utilizing the FERET, Expanded YALE B and ORL repositories. DR-LDP beat the other local descriptor form with a higher identification score of 97.62 percent.

In addition, [18] suggested a facial recognition method focused on PCA, which was introduced using the principle of neural networks. The system's operating theory begins as follows: build a database of recognized individuals with facial images. Then agree on a training range of M number of images corresponding to the variation in facial expressions and lighting conditions of each person. Next, calculate (M × M) matrix (L) and corresponding eigenvectors and its eigenvalues. Then, fuses uniform image training set that generates M Eigen-faces and saves the corresponding values after fusing the image training set together. In addition, the program measures and stores a function vector for anyone in the database to create a different neural network designed for the face of each person found in the facial database. When Eigen-faces are collected, the corresponding computation is performed to acquire feature vectors for the facial images in the database and is given as feedback for increasing neural network training. The training method uses the facial features the same specific person's vectors that are used to train the neural network of an entity and even other neural networks. Once an input image for the identification system is provided then the resulting attribute vectors are determined using already specified Eigen-faces and the new input image representation is retrieved. The ORL face image repository has been used to test the device to demonstrate fair identification rates of 93 percent.

The author in [19] builds an automated method for identifying neutral faces in identification images utilizing deep learning algorithms, and torching the hardware necessary to efficiently execute the established environment for learning consists of 64 GB of RAM and strip-based storage unit. Free CV (open source computer vision), python and Ubuntu (Linux operating system) version 17.10 and Nvidia CUDA 8.0 are the necessary applications. This method was developed using a dataset containing approximately 94 images, the dataset was generated utilizing 128-d embedding for each face in the dataset, the embedding was used to identify the facial pictures characters. After the dataset and folder structure was set, the faces in our training set were quantified using 128 embedding. During classification, the k-NN model was utilized for the final face classification. The system achieved an accuracy of about 95%. It was able to recognize and display the names and face of people in an image. The system can recognize a face image included in a dataset that has been trained.

The study in [20] suggested a Retinex Adaptive Attenuation Quantification (AAQR) approach to improve the overnight image information. This approach contains 3 stages: the constraint of attenuation, the estimation of attenuation and the quantification of adaptations. The efficiency of the proposed model was assessed using a reliable face recognition system via sparse depiction. At night the captured driver's face images were grouped into three categories (UP-Down, Left-Right and Mixed) according to the arrangement of

illumination for each image. The findings revealed that the image recognition levels improved by the suggested AAQR system were 82%, 84%, and 91% respectively for the Up-Down, Left-Right, and Mixed Illumination classes. The detection range of the AAQR system was (2 – 36) percent higher relative to other form of picture improvement. The developed system of successful face recognition focused on the concept element interpretation, genetic algorithm and vector supporting system, in which the key aspect analysis is used to minimize the attribute aspect, the genetic algorithm is utilized to refine the searching technique, and the assistance of the vector system is employed to recognize classification. Through the 2003 simulation study on the face database of the Chinese Academy of Science Institute of Technology, the findings indicate that the design can achieve a higher-efficiency facial recognition and the maximum accuracy rating of 99%.

III. Proposed System

This section provides the framework of the proposed system with its components firstly. Then, it describes the most important component, which is the FR-based system from the Artificial Intelligence (AI) perspective. Finally, it presents the details of building of the FR-based system.

A. Framework of Proposed System

The framework of the proposed system consists of three main components, which as camera, the ready FR-based system, and the data base. The camera is used for face capturing, while the FR-based system takes the image of the face and processes it to be matched with the information stored in the data base. Fig. 4 shows the components of the framework.

As shown in Fig. 4, the login process will be legal if the FR-based system correctly identifies the user as an authenticated one by his face. Otherwise, the system will deny the login process. It is worth mentioning that in Fig. 4 the FR-based system is considered complete and ready for use. However, the process of building the FR-based system is described below.

B. FR-based System in Terms of AI

We can imagine that the system included in the framework described above is delivered to a company to be used by its employees. The employee is allowed to login to his or her office only if the system recognized his or her as authenticated user. In this context and in reality, the delivered system has to be built at the programmers' side and then is used at the company side. According to the fundamentals of the AI, the process of building the FR-based system goes through two main steps. The two steps are illustrated in Fig. 5.

As shown in Fig. 5, there are two main stages in the construction step, which are training and testing. In the training stage, the machine is learnt about how to recognize faces, while in the testing stage, the FR-based system is evaluated in terms of accuracy. In the usage step, a new record (face) is provided as an input to the FR-based system to test the ability of recognition (i.e., ability of dealing and handling new faces that did not train on them previously or did not stored in the data base).

C. Model Construction Step (Training Stage)

The final goal of the training stage is to train the machine to be able to recognize faces. The word "training" means that there must be a database that is used for training purpose, which in general called raw material. In this work, the raw material is represented by the database of faces. Fig. 6 shows the steps of the training stage, where the first step is to select or determine the database.

1) First step: selecting dataset: In this work, the dataset that is used for training is called ORL and obtained from [21]. The dataset contains 400 images of different faces. The images of faces belong to 40 class. The faces included in the dataset vary from persons that wear glasses to persons that has some expressions in their faces. In addition, the images of faces are taken from different angels of light. Moreover, the images are from the size 92 × 112 pixel. Fig. 7 shows the selecting data base step according to the interfaced of the system.

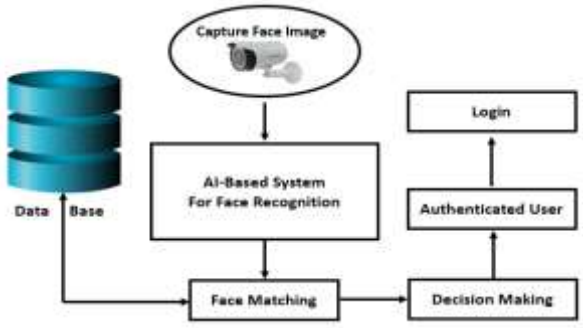

Fig. 4. Components of Framework.

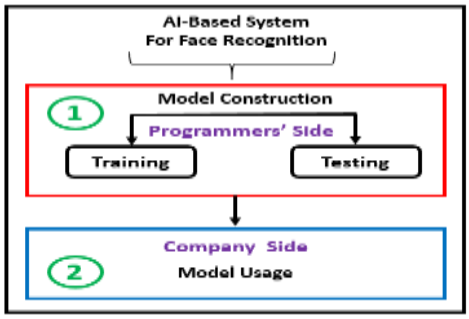

Fig. 5. Steps of Construction and usage of the System.

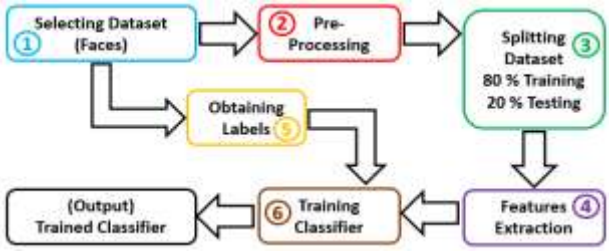

Fig. 6. Steps of Training Stage.

Fig. 7. Selecting Data base with Sample of Image.

Table I shows six images as a sample taken from the used dataset.

2) Second step: pre-processing: The objective of this step is to remove the noise from the image and to crop the face of the person, as shown in Fig. 8.

In reality, the data is noisy. Therefore, removing the noise is essential to prepare the images for training phase. In other words, the classifier will train on clean data, which in turn increases the accuracy rate. Fig. 9 shows the image shown in Fig. 7 after noise removing.

As for the technique used for noise removal, Adaptive Median Filter (AMF) is employed for this purpose. AMF contribute by adding enhancement for the mammogram input images. That is because they have the following benefits [22]:

1) Removal of salt and-pepper (impulse) noise.
2) Smoothing of other noise (may not be impulsive).
3) Reduction of distortion, such as excessive thinning or thickening of object boundaries.

Cropping process means that the face of the person located in the input image will be surrounded by a red rectangle. This in turn means that the Region of Interest (RoI) is accurately determined for further manipulation. Fig. 10 illustrates the RoI for the image used in Fig. 10.

TABLE I. SAMPLE OF IMAGES FROM DATASET

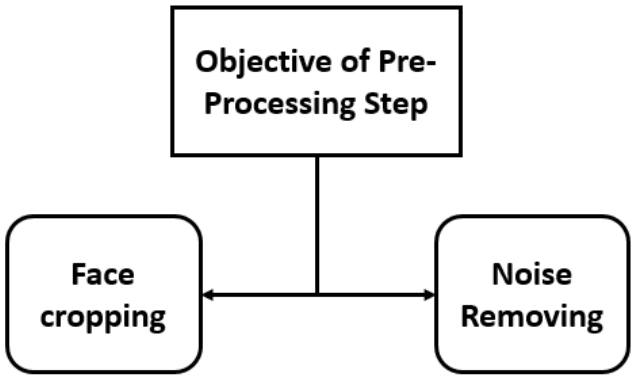

Fig. 8. Objective of Pre-Processing Step.

Fig. 9. Removing the Noise.

Fig. 10. Region of Interest (RoI).

1) Third step: splitting dataset: In this step, the original database is divided into two data sets, which are training data set and testing dataset, as shown in Fig. 11.

As shown in Fig. 11, the training dataset forms 70 % from the original data base, while the testing data set forms 30 % of the original data base. The process of splitting is performed randomly.

Fig. 11. Splitting Database.

2) Forth step: features extraction: In this work, two methods are used for feature extraction, which are Convolutional Neural Network (CNN) and Scale Invariant Feature Transform (SIFT), as shown in Fig. 12.

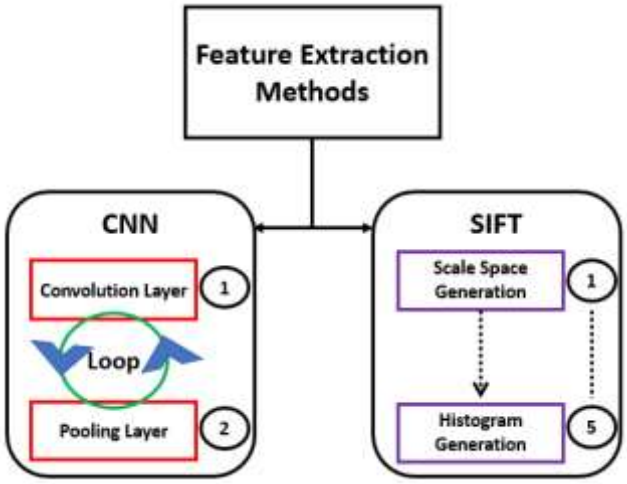

Fig. 12. Methods of Features Extraction.

As shown in Fig. 12, there are two main steps in the CNN method, while there are five steps in the SIFT method. Below in a detailed description of each method.

D. CNN based Method

The method of extracting the features depends on a loop between two main layers in the CNN, which are convolutional layers and pooling layers.

The goal of the convolution layers is to extract simple features from the input image. The goal of the pooling layers is to gather the simple features to from complete and clear features. Fig. 13 illustrates the structure of the CNN with both the convolution and pooling layers.

As shown in Fig. 13, a filter is used for feature extraction. The filter moves in a convolutional manner to scan the whole input image. The convolutional motion leads to generate the features (illustrated by the one-row tables). After features extraction, the pooling process is performed to gather the extracted features to form the final features. It is worth mentioning that the final extracted features are used for training the classifier as described in the 6th step.

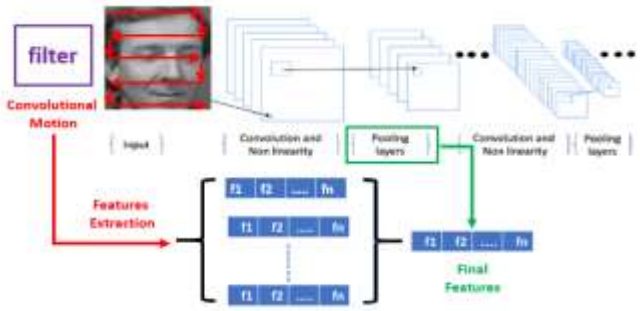

Fig. 13. Structure of CNN in Terms of Convolution and Pooling.

E. SIFT based Method

The objective of the SIFT method [23] is to extraction of features that are stationary even under change in rotation or scale of an image. In general, depending on some interesting points, the rotation invariance is guaranteed. This can be achieved by manipulation both the gradient orientations and the magnitudes of the pixels that are located as a neighbors to the interesting points. As for the scale invariance, it is guaranteed by utilizing a scale space based method. The SIFT has five steps, as illustrated in Fig. 14.

In Fig. 14, the first step is to produce the scale space. This is done by converting the face image through the Gaussian Convolution. This step aims at dealing with images as layers to be an inputs for the next step. The second step is to calculate the difference of Gaussian (DOG). This step is performed by calculating the subtracting of the nearby images. This step aims at facilitating the process of identifying the interesting points. The third step is to determine the most important interesting points. This is done by comparing neighbor pixels with the target pixels in the current and adjacent DOG images. This step aims at facilitating the process of identifying and deleting the poor interesting points (i.e., the points that have low contrast or those that are located in the edges). The forth step is calculating the gradient orientations of the neighbor pixels around the Remained Interesting Points (RIP). This step aims at determining the behavior of the interesting points in terms of directions. The final step of calculating the histogram of the RIP. The histograms are stored in vector for matching process (represented in Fig. 14 by $[a_1. a_2 \dots a_{128}]$).

Compared with the CNN based method, the SIFT based method produces less interesting points in terms of numbers. In other words, the number of interesting points obtained by the SIFT method is less than those obtained by the CNN method. This is due to the filtering process in the third step of the SIFT based method (i.e., removing poor interesting points).

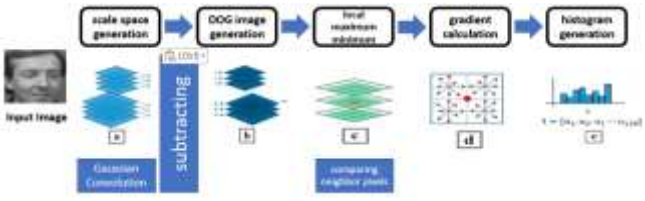

Fig. 14. Steps of SIFT based Feature Extraction Method.

Until now, the previous four steps are performed so that the process of starting training the classifier K-Nearest Neighbor (KNN) is ready. Fig. 15 illustrates that two different methods of features extraction (CNN based method and SIFT based method) are used.

3) Fifth step: obtaining labels: Before starting the process of training the KNN classifier, the labels (classes) of the used data base is obtained. The used data set has 40 class, where each class is denoted by $(S_i|i = 1.2.3 \ldots .40)$. Fig. 16 shows a snapshot of samples of classes from the used data base.

4) Sixth step: training the classifiers: In this step, one classifier is trained on the two types of features that are extracted from both the CNN and the SIFT. The classifier that is used in this work is the KNN. Since there are two methods of feature extraction, we assume that the first classifier that uses the CNN feature extraction method is called C_{CNN}^{KNN}, while the second classifier that uses the SIFT feature extraction method is called C_{SIFT}^{KNN}.

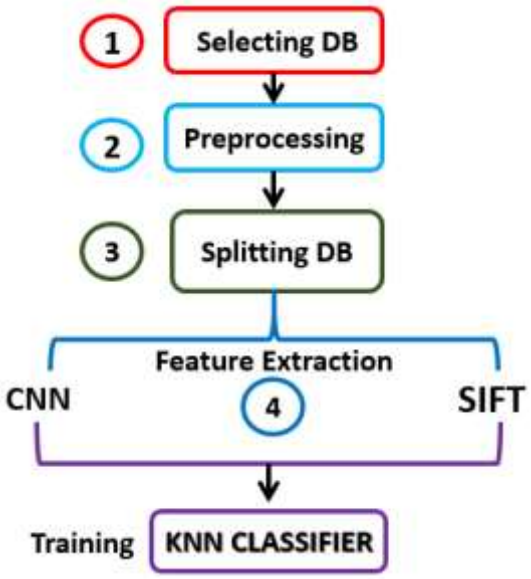

Fig. 15. Flow Chart of the Four Performed Steps.

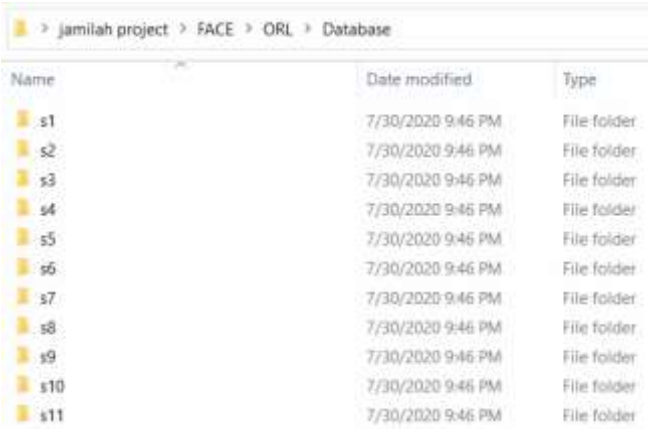

Fig. 16. Samples of Classes from the used Data Base.

F. Training the C_{CNN}^{KNN} Classifier

To train the KNN classifier, an activation function is needed. The activation function that is used in this work is the Softmax function. The reason why the Softmax activation function is used is that it has the following advantages [24]:

1) Able to handle multiple classes only one class in other activation functions normalizes the outputs for each class between 0 and 1, and divides by their sum, giving the probability of the input value being in a specific class.

2) Useful for output neurons, where typically Softmax is used only for the output layer, for neural networks that need to classify inputs into multiple categories.

Visually, the Softmax function is illustrated by Fig. 17.

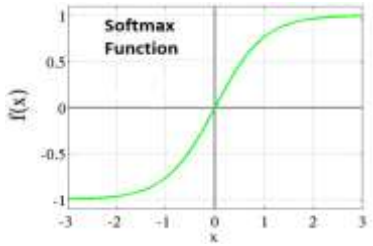

Fig. 17. Softmax Function.

As shown in Fig. 17, the Softmax function is able to represent classes within the range [-1, +1]. Since it has multiple classes property, the 40 classes found in the used data base can be represented. Fig. 18 shows the representation of the 40 classes.

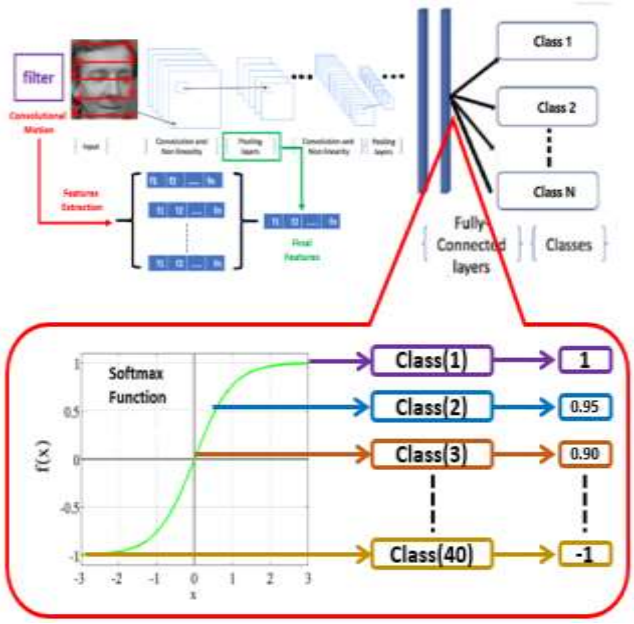

Fig. 18. Using Softmax Function for Classification..

As shown in Fig. 18, the fully connected network is formed (i.e., using the whole features extracted by the CNN). Then, the Softmax function represents the 40 class. In details, since the value that is generated by the Softmax function is limited between the -1 value and the +1 value, there is 2

degrees. To represent the 40 classes numerically, the 2 degrees is divided by 40 to calculate the step of increasing (decreasing).

$$step\ of\ increasing = \frac{2}{40} = 0.05 \quad (1)$$

Depending on the step of increasing, the first class is numerically represented by (+1) value. The 2nd class is numerically represented by (+0.95) value. The third class is numerically represented by (+0.90) value, and so on until the 40th class which represented by the (-1) value.

The KNN classifier works depending on calculating the distance between the features of given image and the center of each cluster. In other words, there will be 40 clusters. Each cluster has a center, which is represented by the value of activation function. For a given face image, the features are extracted, the value of activation function is calculated, and the distance between the value of activation function and each cluster center is determined, and finally the image is assigned to the nearest cluster. Fig. 19 shows an example for the KNN classifier.

Fig. 19 shows the CNN-KNN classifier, where the value of activation function is (-0.98). The centers of the clusters are (+1) for cluster 1, (+0.95) for cluster 2, and so on. The value (-0.98) is closer to the last cluster (its center is -1). Therefore, the image is assigned to the cluster 40.

G. Training the C_{SIFT}^{KNN} Classifier

The process of training the C_{SIFT}^{KNN} classifier is similar to the process of training the C_{CNN}^{KNN} classifier. The difference is related to using histograms as centres of clusters. Consequently, the calculating of distances is conducted between (the histogram of the face image that is under classification) and (the histograms of clusters' centres). Fig. 20 illustrates the process of training the C_{SIFT}^{KNN} classifier.

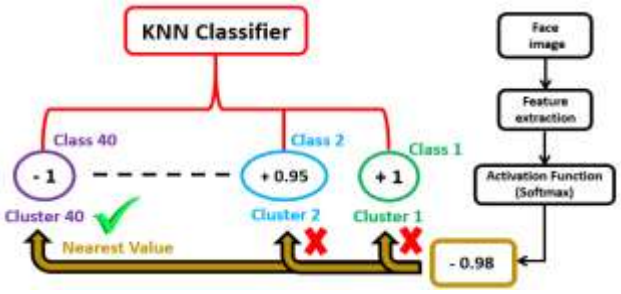

Fig. 19. CNN-KNN Classifier.

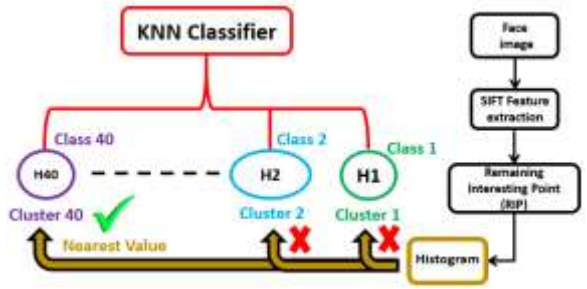

Fig. 20. SIFT-KNN Classifier.

IV. Metrics for Evaluation

To evaluate (test) the built two classifiers, the testing set is used. The built classifiers either classify a given face image correctly or incorrectly. The testing set contains 120 face images, as shown in Fig. 21.

Fig. 21. Testing the Classifiers.

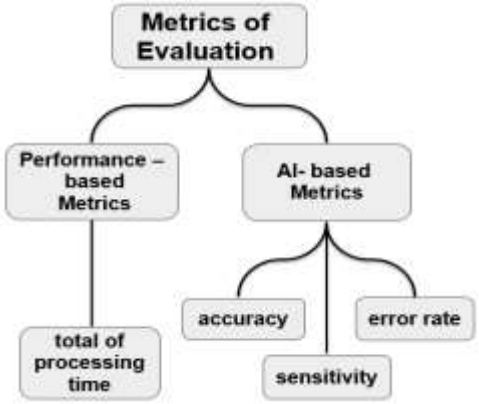

Fig. 22. Evaluation Metrics.

Two kinds of metrics are employed in the process of evaluation the proposed model. They are the AI-based metrics and the performance based metrics, as shown in Fig. 22.

In general, a confusion matrix is an effective benchmark for analyzing how well a classifier can recognize face images of different classes [25]. The confusion matrix is formed based on the following terms:

1) True positives (TP): Positive records that are correctly labelled by the classifier.

2) True negatives (TN): Negative records that are correctly labelled by the classifier.

3) False positives (FP): Negative records that are incorrectly labelled positive.

4) False negatives (FN): Positive records that are mislabeled negative.

Table II shows the confusion matrix in terms of the TP, FN, FP, and TN values.

TABLE II. Confusion Matrix

Actual class (Predicted class)	**Confusion matrix**		
	C1	**¬ C1**	**Total**
C1	True positives (TP)	False negatives (FN)	TP + FN = P
¬ C1	False positives (FP)	True negatives (TN)	FP + TN = N

Relying on the confusion matrix, the accuracy, sensitivity, and error rate metrics are derived. For a given classifier, the accuracy can be calculated by considering the recognition rate, which is the percentage of face images in the test set that are correctly classified. The accuracy is defined as:

$$Accuracy = \frac{(TP+TN)}{number\ of\ all\ images\ in\ the\ testing\ set\ (120)} \quad (2)$$

Mechanisms for accuracy-based evaluation. In this context, a higher accuracy corresponds to a better classifier output. The maximum value of the accuracy metric is 1 (or 100%), which is achieved when the classifier classifies the all face images correctly without any errors in the classification process.

Sensitivity refers to the true positive recognition rate. It is given by:

$$Sensitivity = \frac{TP}{P} \quad (3)$$

Mechanisms for sensitivity-based evaluation. In this context, a higher sensitivity corresponds to a better classifier output. The maximum value of the sensitivity metric is 1 (or 100%), which is achieved when the proportion of true positive cases equals the number of actual positive cases.

The error rate is defined as the ratio of mistakes made by the classifier during the prediction process. It is defined as:

$$eror\ rate = 1 - accuracy \quad (4)$$

Mechanisms for error rate-based evaluation. In this context, a lower error rate corresponds to a better classifier output. The minimum value of the error rate metric is 0, which is achieved when the classifier classifies all the records correctly (i.e., the accuracy is 100%).

Time dominates the situation when it comes to talking about performance metrics. In other words, the total time of stages (T_{stages}^{time}) required to build the classifier is used as a benchmark. The T_{stages}^{time} is given by:

$$T_{stages}^{time} = T_{stage}^{prep} + T_{stage}^{splittingDB} + T_{stage}^{FeEx} + T_{stage}^{OL} + T_{stage}^{Trn} \quad (5)$$

where T_{stage}^{prep} refers to the preprocessing time, $T_{stage}^{splittingDB}$ refers to the database splitting time, T_{stage}^{FeEx} refers to the features extraction time, T_{stage}^{OL} refers to the labels obtaining time, and T_{stage}^{Trn} refers to the training time. It is well known that the shorter the total time is, the higher level of performance.

V. Results and Discussions

This section is organized so that the setup is firstly presented, which describes the environment where the experiments are conducted. Then, the results with corresponding discussions are provided.

A. Setup

The context within which the experiments are conducted is shown in Fig. 23.

As shown in Fig. 23, the Matlab programming language (version: 2018) is used for implementing the proposed face recognition system. After finishing implementation stage, the proposed system is executed on a machine that has (capacity of RAM: 16 GB, and speed of Processor: 2.59 GHz). Both the CNN-based KNN and the SIFT-based KNN classifiers are involved in the comparison.

B. Results

The results are provided in practical style firstly, and then in a numerical style for more analyzing and discussion.

1) Practical style of results: The results are provided through an execution of the program, showing the results of classification in both the CNN-based KNN classifier and the SIFT-based KNN classifier. Fig. 24 and 25 shows the practical results.

The CNN-based KNN classifier shows better values in terms of accuracy (95.95 %) when compared to the SIFT-based KNN classifier (94.60 %).

2) Numerical style of results: The results are provided through the values of the AI-based metrics and the performance based metrics.

To obtain the values of the AI-based metrics, it is required to execute the program many several times. Since the testing data set contains (130 face images), it is required to execute the program 130 times using the CNN-based KNN classifier, and the same executions (on the same 130 face images) are then repeated using the SIFT-based KNN classifier. Fig. 26 illustrates the mechanism of obtaining the accuracy results.

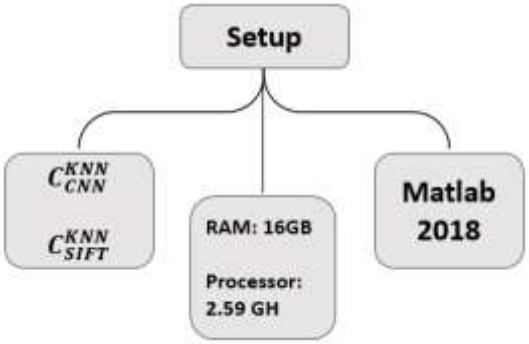

Fig. 23. Setup of Experiments' Environment.

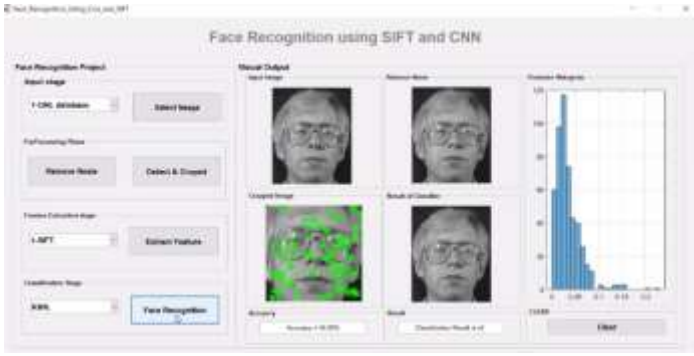

Fig. 24. Prediction of Face Class using the SIFT-based KNN Classifier.

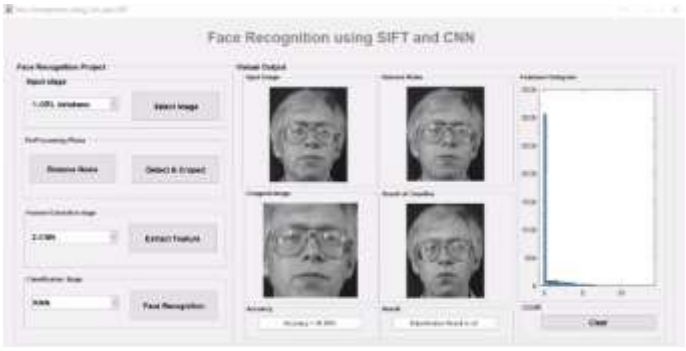

Fig. 25. Prediction of Face Class using the CNN-based KNN Classifier.

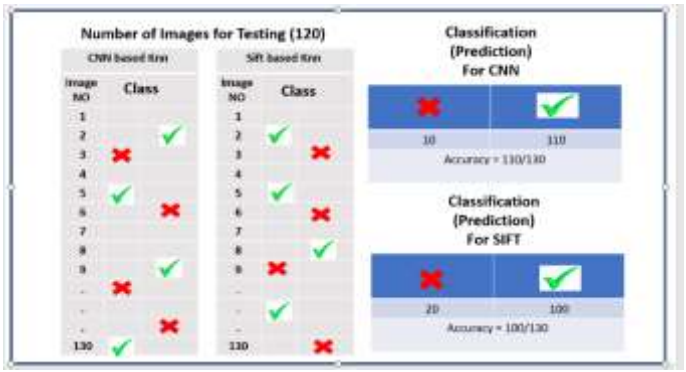

Fig. 26. Mechanism of Obtaining the Accuracy Results.

Table III shows the values of the AI-based metrics after executing the face regression system (270 time) according to the mechanism illustrated by Fig. 26.

Discussion. According to the results arranged in Table III, the CNN-based KNN classifier classified 116 face images correctly, while it misclassified 4 face images. The SIFT-based KNN classifier classified 114 face images correctly, while it misclassified 6 face images. The reason behind these results is related to the concept of both the CNN and the SIFT. In other words, the number of features that are extracted by the CNN is more than the Features that are extracted by the SIFT since the latter has a filtering phase. In addition, the CNN scans the whole face image, while the Sift scans only the space of scale that is determined. The values of the error rates support the results obtained under the accuracy metric. As for the sensitivity, the CNN-based classifier shows better values when compared to the SIFT-based classifier. That is because the sensitivity is related to how many images that are classified as true, and this increases with the increasing of the accuracy.

Fig. 27 shows the results depending on the performance-based metrics.

TABLE III. COMPARISON OF CLASSIFIERS

Classifier	**Metrics**	
	Accuracy	***Error rate***
CNN-based	97 %	3 %
SIFT-based	95 %	5 %

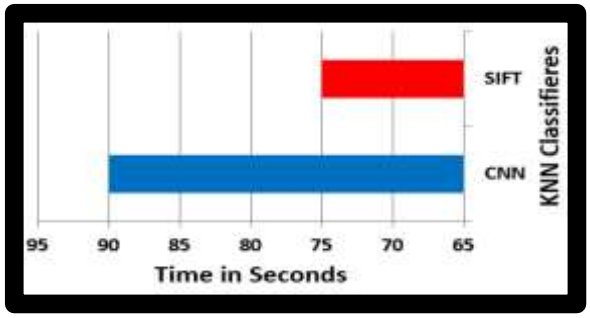

Fig. 27. Performance of the Two Classifiers.

Discussion. As shown in Fig. 27, the performance of the CNN-based KNN classifier is less than the performance of the SIFT-based KNN classifier. That is because the time required to execute the SIFT feature extraction method is shorter than the time required to execute the CNN feature extraction method. The reason behind this is related to the process of determining specific points of interest in the SIFT method. In the CNN method, the whole face image is scanned and the whole features are extracted and not be limited in specific points of interest. However, although the SIFT outperformances the CNN, the CNN achieves higher level of accuracy. In terms of cyber security, the accuracy level is preferred since the system must has the highest ability to recognize authorized users.

VI. CONCLUSION

Ensuring high level of authentication is a critical issue in cyber security systems. AI can be used to build strong face recognition systems that have the ability of providing high level of authentication as a required cyber security requirement. However, the system is poor if the degree of the accuracy is low. In this work, the CNN-based KNN and the SIFT-based KNN classifiers are proposed for face recognition. The CNN-based classifier uses the CNN itself to extract the features without using the last layers that are responsible for classification. The SIFT-based classifier uses the standard five steps of the SIFT method for feature extraction. A standard data base (ORL) is used for training and testing the classifiers. The two proposed classifiers are compared according to the accuracy, sensitivity, error rate, and time of response. The CNN-based classifier showed better results according to the AI-based metrics.

Limitation: This work did not take into consideration the performance of the CNN-based classifier. It is considered put of scope in this work since it is related to achieving high level of security.

Future work: In future work, we intend to enhance this work in terms of performance by employing Hadoop platform. In addition, the privacy and other security requirements will be taken into account.

REFERENCES

[1] Baker, Nathan, et al.Workshop report on basic research needs for scientific machine learning: Core technologies for artificial intelligence. USDOE Office of Science (SC), Washington, DC (United States), 2019.

[2] Nixon, Mark, and Alberto Aguado. Feature extraction and image processing for computer vision. Academic press, 2019.

[3] Aileni, Raluca Maria, et al. "Cybersecurity Technologies for the Internet of Medical Wearable Devices (IoMWD)." Advances in Cyber Security Analytics and Decision Systems. Springer, Cham, 2020. 117-140.

[4] Sajjad, Muhammad, et al. "Raspberry Pi assisted face recognition framework for enhanced law-enforcement services in smart cities." Future Generation Computer Systems 108 (2020): 995-1007.

[5] Kumar, Priyan Malarvizhi, et al. "Intelligent face recognition and navigation system using neural learning for smart security in Internet of Things." Cluster Computing 22.4 (2019): 7733-7744.

[6] Revate, S. S. "Is Biometric Security System Safe from Viral Infection Covid19?." Purakala with ISSN 0971-2143 is an UGC CARE Journal 31.9 (2020): 181-189.

[7] Basin, David, et al. "A formal analysis of 5G authentication." Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. 2018.

[8] Kant, Krishna. "Data center evolution: A tutorial on state of the art, issues, and challenges." Computer Networks 53.17 (2009): 2939-2965.

[9] Kemelmacher-Shlizerman, Ira, et al. "The megaface benchmark: 1 million faces for recognition at scale." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[10] Abdullah, A. S., Abed, M. A., & Al Barazanchi, I. (2019). Improving face recognition by elman neural network using curvelet transform and HSI color space. Periodicals of Engineering and Natural Sciences, 7(2), 430-437.

[11] Khan, M. Z., Harous, S., Hassan, S. U., Khan, M. U. G., Iqbal, R., & Mumtaz, S. (2019). Deep unified model for face recognition based on convolution neural network and edge computing. IEEE Access, 7, 72622-72633.

[12] Rejeesh, M. R. (2019). Interest point based face recognition using adaptive neuro fuzzy inference system. Multimedia Tools and Applications, 78(16), 22691-22710.

[13] Gupta, S., Thakur, K., & Kumar, M. (2020). 2D-human face recognition using SIFT and SURF descriptors of face's feature regions. The Visual Computer, 1-10.

[14] M. R. Faraji and X. Qi, "Face recognition under varying illumination with logarithmic fractal analysis," IEEE Signal Process. Lett., vol. 21, no. 12, pp. 1457–1461, 2014.

[15] B. H. Shekar and D. S. Rajesh, "Affine Normalized Krawtchouk Moments Based Face Recognition," Procedia Comput. Sci., vol. 58, pp. 66–75, 2015.

[16] P. Zhang, X. Ben, W. Jiang, R. Yan, and Y. Zhang, "Coupled marginal discriminant mappings for low-resolution face recognition," Optik (Stuttg)., vol. 126, no. 23, pp. 4352–4357, 2015.

[17] S. P. Ramalingam and P. V. S. S. R. Chandra Mouli, "Robustness of DR-LDP over PCANet for face analysis," Int. J. Multimed. Inf. Retr., vol. 7, no. 2, pp. 129–137, 2018.

[18] A. Mandhare and S. Kadam, Performance Analysis of Trust-Based. Springer Singapore, 2019.

[19] P. C. Okoye and E. A. Adenagbe, "Development of a Face Recognition System with Deep Learning and Pytorch," Int. Res. J. Eng. Technol., vol. 6, no. June, pp. 3439–3441, 2019.

[20] J. Shen et al., "Nighttime driving safety improvement via image enhancement for driver face detection," IEEE Access, vol. 6, no. c, pp. 45625–45634, 2018.

[21] kaggle(websit,2020.[line]available; https://www.kaggle.com/kasikrit/att-database-of-faces access 12septomper 2020.

[22] Saleem, S. Abdul, and T. Abdul Razak. "An effective noise adaptive median filter for removing high density impulse noises in color images." International Journal of Electrical and Computer Engineering 6.2 (2016): 611.

[23] Alberry, H. A., Hegazy, A. A., & Salama, G. I. (2018). A fast SIFT based method for copy move forgery detection. Future Computing and Informatics Journal, 3(2), 159-165.

[24] Mona Alfifi, Mohamad Shady Alrahhal, Samir Bataineh and Mohammad Mezher, "Enhanced Artificial Intelligence System for Diagnosing and Predicting Breast Cancer using Deep Learning" International Journal of Advanced Computer Science and Applications(IJACSA), 11(7), 2020. http://dx.doi.org/10.14569/IJACSA.2020.0110763.

[25] Luque, A., Carrasco, A., Martín, A., & de las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. Pattern Recognition, 91, 216-231.