International Journal of Future Engineering Innovations

# A Comparative Study of Face Recognition and Detection Mechanisms Through deep and Machine Learning and Handcrafted Features

**Lana Abdullah AL-Afeef [1], Hazem Munawer Al-Otum [2]**
[1] Master of Electrical Engineering, Department of Electrical Engineering, Jordan University of Science & Technology, Jordan
[2] Prof., Ph.D. EE Department, Faculty of Engineering, Jordan University of Science & Technology, Jordan

* Corresponding Author: **Lana Abdullah AL-Afeef**

## Article Info

**ISSN (online):** 3049-1215
**Impact Factor (RSIF):** 8.25
**Volume:** 03
**Issue:** 04
**Received:** 09-05-2026
**Accepted:** 08-06-2026
**Published:** 07-07-2026
**Page No:** 52-64

**Abstract**
In this paper, a comparative study between handcrafted and automated feature extraction method has been provided. The handcrafted method has been based over local binary pattern (LBP) as feature extraction technique. The histogram equalization (HE), multi-scale retinex (MSR), and a difference of Gaussian (DOG) have been used as a preprocessing technique to improve the image quality. The results of the handcrafted approach have been shown that the performance with HE is the best. In the automated part, ALEXNET has been used as convolutional neural network (CNN) architecture. The standard gradient descent with momentum (SGDM) has been used as the optimizer, because the results were better when it has been used. The results of the automated part have been shown how the layers activation functions works. In the automated part, the training and test accuracy have been evaluated and compared between different databases. The accuracy has achieved up to 100% in Face94 and Face grimace databases as the best accuracy in the CNN approach.

**DOI:** https://doi.org/10.54660/IJFEI.2026.3.4.52-64

**Keywords:** Deep learning, Face recognition, Convolutional neural network, Feature extraction, Local binary pattern

## 1. Introduction

Today the high growth in the internet in social network draws the attention of security professionals. Face recognition is considered one of the most important topics that have many applications. The technologies of face recognition have been widely used in various applications, including biometric authentication, protection, human computer interaction and multimedia management (Trigueros *et al*., 2018). Research into face recognition is not only motivated by specific challenges emerging from a problem of recognition, but also by various practical applications where identification is required. Becoming one of the leading biometric technologies has become increasingly relevant as a result of rapid technological advances such as digital cameras, the Internet, mobile devices and an increased security need. Over other biometric technology, facial recognition has many benefits. It is essentially non-intrusive and easy to use. It may work in two modes, one or both, which are face verification or authentication and identification or recognition of the face. Face verification requires a one-to-one match that compares a face image of the query with a face image of the template in the dataset, so it is possible to check the face image whether it is accurate or not. Face recognition involves a number of matches comparing a query face image with all template images in the dataset to determine the identity of the query face (Crosswhite *et al*., 2018). The task became a difficult endeavor, despite the advancement in face recognition, particularly for the unregulated setting where illumination, expression and poses have varied significantly. Based on the position points, a facial part is placed, such as eyes, nose, mouth, and facial outlines. Features are extracted after the normalization of the face in terms of geometry and metric so that the effective information is obtained that can distinguish between the faces of different people and stable information characterizing the geometrical and photometric changes. Inputs face vector features that are nearly orthogonal to those of the registered faces in the face matching database. It reports the face identification when the match is with a recognized face and unknown face. The interpretation of face recognition results is highly dependent on features that are extracted to present the face pattern, as well as on classification methods which are used to oppose such features from differentiating the faces. Face localization and normalization provide the basis on which efficient features are extracted. Even if the datasets are different and have different face recognition scenarios, the main objective of evaluation is the same – it is the classification performance. The classification accuracy is the most important consideration of the performance evaluation in different scenarios and is tested on a benchmark (Le, 2011). However, unlike the face identification where any accuracy less than 100% implies errors are being made, the decision here is made in binary whereas the claimed identity is either correct or it is not, which affects the accuracy of the face verification systems. Face verification accuracy indicates how often, on average, a verifier makes the right choice with regard to the provided identity. The verification accuracy, in this case, is usually judged using the Receiver Operational Characteristic curve. This curve shows the relationship between True Acceptance Rate and False Acceptance Rate (Crosswhite *et al*., 2018).

The problem with engineering characteristics is that specialists conduct research using the methods of the reliability of each form to the various variations in an unstructured environment, that is, methods of consistency with age, position, illumination invariant means (Liu *et al*., 2005; Tan & Triggs, 2010; Park *et al*., 2010; Li *et al*., 2011; Ding & Tao, 2016). Additionally, there are several methods of texture feature extraction (Chaki & Dey, 2020). Four main categories are separated. Method statistical characteristics (Ohanian & Dubes, 1992; Varma & Zisserman, 2005; Moraru *et al*., 2019) and methods based on models [13, 14], as well as the results of pixel tribes primarily analyze the spatial relationship around the pixel small. The best known are the local binary image histogram patterns [15, 16, 17, 18, 19]. Regular ways of determining the texture of structures [20] are considered structural methods. Methods for analyzing and classifying textures were developed by many authors using the results of filter-based research. This includes Wavelet transformation [21].

One of the latest developed systems is called Hess- ACS-LBP, a novel hybrid Local Binary Pattern based on Hessian matrix and Attractive Center Symmetric LBP. The Hessian matrix provides the set of directional derivatives for different texture regions, while ACS-LBP effectively shows the local texture features. The application of Hess-ACS LBP leads to the respective increase for the classification accuracy by about 20 percent in the case of the initial LBP system and from 1 to 11 percent concerning other state of the art LBP methods. Thus, for example, in the CUReT and Mondial Marmi databases, the conventional LBP method has 91.03 and 79.06 percent for classification accuracy. for the ACS-LBP approach, the accuracy is 94.72 and 91.99 percent accordingly. Meanwhile, the Hess-ACS-LBP proposed has 97.06% and 100% for classification accuracy for both cases, but its complexity is a disadvantage [22]. A Learned Local Gabor Patterns (LLGP) has been proposed in [23]. The present technique deals with the representation and recognition of the face. The Gabor features are used and the intensity mapped masking is used to reduce the intensity of the context of the face image. The Laplacian Gaussian filter is used which is used for the enhancement of the edges while the BPSO method is used for the optimal selection of the features and the classifier used for this technique is the Euclidean classifier. The traditional methods for face recognition are replaced with the deep learning techniques of coevolutionary neural networks. While learning, these deep learning techniques learn with the very large datasets and these learn the best features to represent the datas as per the learning. Now a day the datas for face is easily available on the internet [25-31] that contain real- world variations. Many CNN architectures have been introduced which is designed in an unconstraint environment [24].

Coevolutionary neural networks have demonstrated an outstanding capacity to learn face images that change according to age, pose, illumination and expression [28-33]. A traditional CNN architecture uses several conveyor strata and the completely connected layer and classifier SoftMax to be connected. CNN's fully-related layer features provide high level semantic information, in opposition to handcraft features [29]. Additionally, faces are easily cropped out of the image or video frame by face recognition algorithms. In a framework developed by Facebook, which they call Deep Face, a 3D facial model is used to apply a piecewise linear texture map to the face and then extracted using a deep 9-layer neural network [30]. Their algorithm was trained on four million faces of 4,000 people, the largest dataset with a clear accuracy of 97.35% classification. As for the current face recognition technology, it is described in [31]. In [32] at the core, we constructed an architecture based on a traditional residual network architecture. The changed network main goal was to identify the face. The recognition results show that the constructed CNN is ahead of the latest technology and shows the best accuracy in publicly available datasets, all types, including the most complex ones: LFW, Face94, Face95, Face96, Grimace. Unlocking software on a cell phone is another program developed and used on the Android market [33]. The owner can be safely authenticated using this unlocking technique. Other algorithms made use of the 3D shapes for the data-level fusion of texture and face recognition [34]. The algorithm was trained and tested using face databases like Face 94, Face 95, Face 96, FERET, and FRGC [35].

The issue of facial recognition could be very broadly interpreted as facial check and facial detection. The system is assessing whether double-hand observations come from the same identity during the authentication; here, a one-to-one resemblance of a probe to a reference photo is calculated. During the identification stage, a single-to-many resemblances is calculated between an image source and a medium of registered criteria to determine how many camera subjects. In particular, when it comes to face authentication, we need to ensure that the person in question is who they say they are or whom we think they are for access control or other verification. Facial identification is critical for record review or evidence hunt for watchlist surveillance. The problem of face check is typical for face-recognition reviews. The last five decades have seen a tremendous spike in the amount of the human beings and their facial images as well as the complexity of the raw data in various databases that has reportedly been improperly prepared [36]. Face recognition is considered the most prevalent and essential field out of all biometric traits. The Face Recognition Grand Challenge data base conducted exhaustive studies (FRGC) [37]. Similarly, the YouTube Faces dataset pairs of subject videos for research [38]. These data sets have set the benchmark for face recognition, with increasing results [39-42]. In this paper, we provide a comparative analysis between hand c rafted and automated face recognition techniques. The paper proposes a handcrafted face recognition system that implements the use of local binary pattern as feature extraction techniques with pre-processing techniques difference of gaussian, Multi scale retinex and Histogram equalization to solve one of the problems in fa ce images, due to real capture system i.e. lighting variations. The a utomated face recognition system is implemented using ALEXNET architecture over different databases. The remainder of the paper is organized as follows This paper is organized as follows. Section 2 describes the materials and methods used to evaluate the performance of the face recognition. Section 3 describes the experimental setup and the results. Finally, Section 4 presents our conclusions.

## Research Method

### Handcrafted Feature Extraction

The intricately designed handcrafted features necessitated judicious balancing between precision and computational efficiency. Over time, the expanded repertoire of features better addressed diverse challenges, as described in related works [43]. Central to the face verification system's functionality were multiple determinants, including illumination variations presenting stubborn obstacles [44].

Preprocessing approaches aimed to remedy pervasive issues in facial imaging databases involving lighting and illumination. As Figure 1's schematic outlines, one of three filters preprocessed the visage, whose attributes the LBP extractor then derived. Performance evaluations using ROC curves followed. Meanwhile, others straightforwardly depicted the block diagram and feature extraction process. To evaluate the performance the ROC curve has been used. The preprocessing technique is important for enhancing image information that removes undesired distortions and improves certain image features that are appropriate for further processing and analysis tasks. When an image is acquired by a camera, often these images cannot be used directly. Random variations in intensity, variations in lighting, or poor contrast that must be resolved in the early stages of visual processing can corrupt the image. This section addresses image enhancement methods to remove these undesirable features. This section begins with a difference of Gaussian (DOG), followed by a brief review of multi-scale retinex (MSR), and then histogram equalization (HE). In the next subsections, a detailed discussion will be provided for each of the block in figure 1.

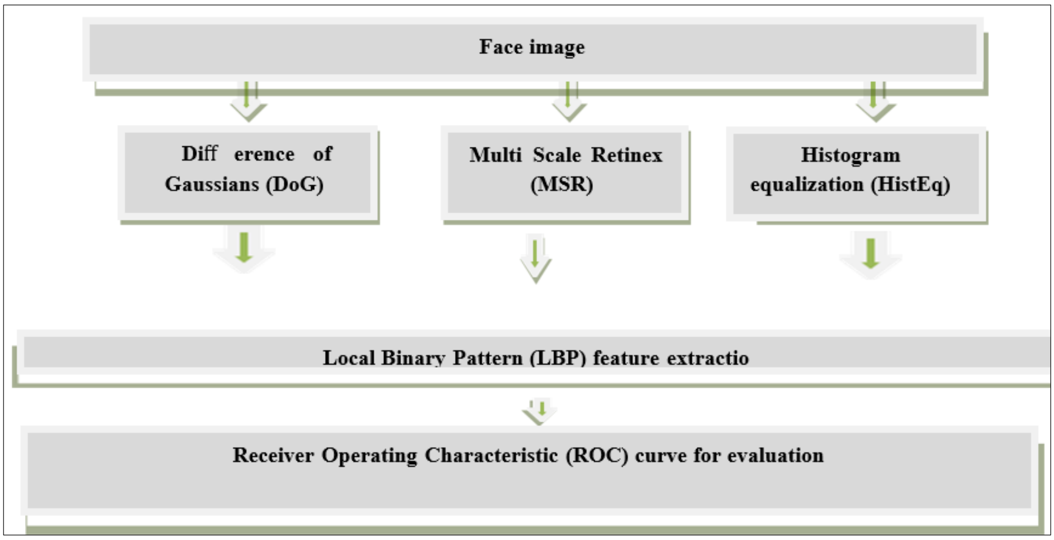


**Fig 1**: General Handcraft Feature Extraction Block Diagram

### Difference of Gaussian (DOG) Algorithm

DoG is a grayscale image enhancement algorithm that effectively eliminates unwanted shadowing effects. The DoG difference is calculated as subtracting a Gaussian blurred version of an image from another less blurred rendition [45]. Namely, the transformation involves deducting an extremely diffused rendition from a less obscured version to function as a band-pass filter maintaining a particular spatial frequency. DoG filters remove spatial elements of high frequency reflecting noise in the image. Through the distinctive process, the noise is extracted via blurring while low frequency components symbolizing uniform areas in the image are maintained. Alternatively, shorter sections of text could be complemented by more convoluted sentences to enhance burstiness, balancing complexity across the passage [57]. Basically, it is a band pass filter capable of suppressing the variations of lighting in the low band and, at the same time, the high band noise. Displaying the contours is very effective. The DoG of Image I is obtained in one dimension by subtracting the image which the researcher has combined with the Gaussian of variance from an image which combined with a Gaussian of smaller variance as follow [46]:

Displaying the contours is very effective. The DoG of Image I is obtained in one dimension by subtracting the image which the researcher has combined with the Gaussian of variance from an image which combined with a Gaussian of smaller variance as follow [46]:

Displaying the contours is very effective. The DoG of Image I is obtained in one dimension by subtracting the image which the researcher has combined with the Gaussian of variance from an image which combined with a Gaussian of smaller variance as follow [46]:

$$\Gamma_{\sigma,K\sigma}(x,y) = I * \left(\frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{(2\sigma^2)}} - \frac{1}{2\pi K^2\sigma^2} e^{-\frac{(x^2+y^2)}{(2K^2\sigma^2)}}\right) \quad (1)$$

Where $\Gamma_\sigma 1$, $k\sigma 2$ (x, $y$) is the one dimensional difference of Gaussian (DOG) of the image, I is the input image,$\sigma_1$, $\sigma_2$ are the standard deviation of a Gaussian filter,$\sigma^2$, $\sigma^2$ are thevariance of a Gaussian filter where $\sigma_2 > \sigma_1$ in other word $\sigma_2$ is bigger than $\sigma_1$ by a value k.

**Multi Scale Retinex (MSR) Algorithm**

The MSR Algorithm is an enhancement of the single space retinex, a member of the surround function class that determines the deviation between the input value and its circular average center. The Multi Scale Retinex algorithm attempts to solve the problem of halos. Halos are bands of light that run along the edges of an image, and may appear bright or dark. The MSR uses three Gaussian filters, hence using three different bandwidths. In other words, as indicated in, it is a weighted sum of three SSR. MSR can be presented is as [47]:

$$R_{MSRi} = \sum_{n=1}^{N} Wn\, R_{n1} = \sum_{n=1}^{N} Wn[\log Ii(x,y) - \log(Fn(x,y) * Ii(x,y))] \quad (2)$$

Where N is the number of scales, wn is the weight of every nth scale, Rni is the ith component of the nth scale, RMSRI is the MSR-output ith spectrum component and Ii(x,y) is the image distribution in the ith spectral band. The notation "*" here defines the convolution function and the surround function can be defined as Fn(x,y) [47]:

$$\mathrm{Fn}(x,y) = C_n \exp[-\frac{(x^2+y^2)}{2\sigma^2}] \quad (3)$$

Where σ is the standard filter deviation, that monitor the amount of spatial detail which is kept and C ensures the normalizing factor [58]:

$$\int F(x,y)dxdy = 1 \quad (4)$$

However, the immediate questions are a) which scales to use, b) and how many, and c) what the weights are. According to Jobson *et al*., three scales are appropriate to most pictures and that the weights may be equal. The scales were set on 15, 80 and 250 experimentally by Jobson *et al*. [47]. The constants of Multiscale Retinex in the software implementation are listed in table 1, where, G and b are the final gain and offset values, respectively, α is a nonlinearity strength controlling constant, and β is a constant of gain value in the color restoration function.

**Table 1:** List of constants in the implementation of MultiscaleRetinex [64]

| Constant | N | $\sigma_1$ | $\sigma_2$ | $\sigma_3$ | $\alpha$ | $\beta$ | $w_n$ | G | b |
|---|---|---|---|---|---|---|---|---|---|
| value | 3 | 15 | 80 | 250 | 125 | 46 | 1/3 | 192 | -30 |

**Histogram Equalization (HE)**

The histogram equalization technique (HE) is the most common way of changing the image intensity to increase the poor image quality. HE technique can be implemented in biometrics textures such as fingerprints, medical images [48], and segmentation of textures [49, 50].

However, the histogram of an image is mainly the comparative frequency of occurrence in the image of the different gray levels with gray levels in the range [0, L-1] which is given by [51]:

$$P_{n} = \frac{\text{Number of pixels with intensity n}}{\text{Total number of pixels}} \quad n = 0, 1, \ldots, L-1 \quad (5)$$

Where, P denotes the normalized histogram for each possible intensity, L is the number of possible intensity values, n is an intensities that ranging from 0 to 255 in a [mr *nc] matrix of the image.The histogram equalized image g will be defined by[59]:

$$g_{i,j} = [\,(L-1)\sum_{n=0}^{f\,i,j} Pn\,] \quad (6)$$

Where, the floor ⌊.⌋ that rounds down to the nearest integer.

**Uniform Local Binary Patterns (LBP) Feature Extraction Algorithm**

There are many different techniques to extract the most critical features from p re-processed f acial images to detect faces, and the Local Binary Pattern is one such method. LBP was first described by Ojalaet al. in 1996 and is a relatively new approach [52]. With LBP, the texture and shape of a digital image can be represented. Image histograms, or features, need to be extracted to split an image into many, small image regions. Since the LBP operator scans each measurement area and extracts the most frequently occurring pattern, which eliminates the imprecise differentiation of pixel quantification in the terrain, it creates a feature that characterizes the pixel in the areas. Because this can be the fused feature of the areas, it creates a picture portrait as a histogram of function. The likeness of two images may be calculated by comparing their histograms. The LBP technique is one of the best methods for face pictures, both in terms of speed and discrimination accuracy [53, 54, 55]. Figure 2 shows, the approach seems very resistant to facial expressions, varying light conditions, diversity of the skin hue, the rotation of the image and the aging of the individual. The LBP operator uses the value of the pixel center as a threshold with eight neighbors of a pixel. If the value of the nearest pixel has a gray value higher than the pixels center or the same gray value these computations for this specific distance between pixels is given the pixel is assigned one, and if the value is less, then it is given zero. Looking back at the image that represents the LBP operator, we can generate a binary code representing the central pixel seen above as a center.

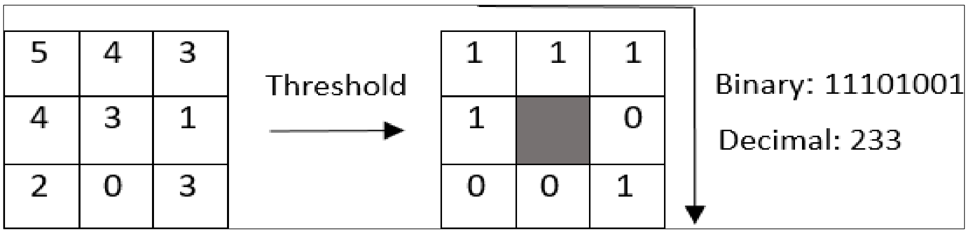


**Fig 2:** The Original LBP Operator

The LBP operator was later expanded to use neighborhood of varying sizes. In this case, the radius R of the center pixel is made a circle. In order to get the values of all the neighborhood sampling points for any radius and any number of pixels, P sampling points on the edge of this circle are taken and compared with the value of the center pixel. For neighborhoods, the notation (P, R) is used. This implies that, for the next point that has a greater gray value (or the same value) of the center pixel, one will be assigned to that point, otherwise it will be assigned a zero as follow [56]:

$$s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \quad (7)$$

Where x is the pixel value and $s(x)$ is the new value of the pixel after the LBP comparison process. The LBP can produce for a center pixel and is described by [56]

$$LBP_{P.R}(x_{c,}y_c) = \sum P - 1 \sum_{P=0}^{P=1} s\ (g_p - g_c)\ 2^P \quad (8)$$

Where (xc, yc) is the coordinates of the center pixel, $gc$is the gray value of the center pixel, $gP$ is the gray value of the center pixel with P = 0,.., P – 1. So, when using (8, R) neighborhood, there are a total of 256 patterns 58 of which are uniform, which yields in 59 diff erent labels. Figure 3 shows An example of circularly Neighboring groups for three different values of P and R.

A Local Binary Pattern is called uniform if it comprises at most two bitwise transitions from 0 to 1 or vice versa. In other words, this means there either uniform pattern with zero transitions or uniform pattern with two transitions. With one transition, it is impossible, because you should think of the binary ordered set as rotating perpetually. The two versions are 00000000 and 1111111 with zero transition, for example eight bits. Examples are 00011100 and 11100001 with a uniform pattern of eight bits and two transitions. The only uniform LBP has two big advantages. The first is that it saves memory. The second is that it only picks up one feature, a uniform LBP, such as spots, line ends, edges and corners. These advantageous features are the ones that contain considerable local textures. There are possible LBPs with LBP. The LBP process can also be applied on images (of faces) to extract features, which can be used to calculate the similarity between these images. The principle of a verification process with face recognition is shown in figure 4. The verification process can be broken into a face representation phase and a feature extraction phase. To represent the input face, it can be thought of as a compound of micro-patterns that the LBP operator can effectively detect, as shown in figure 4. The input face is best represented as an LBP feature histogram.

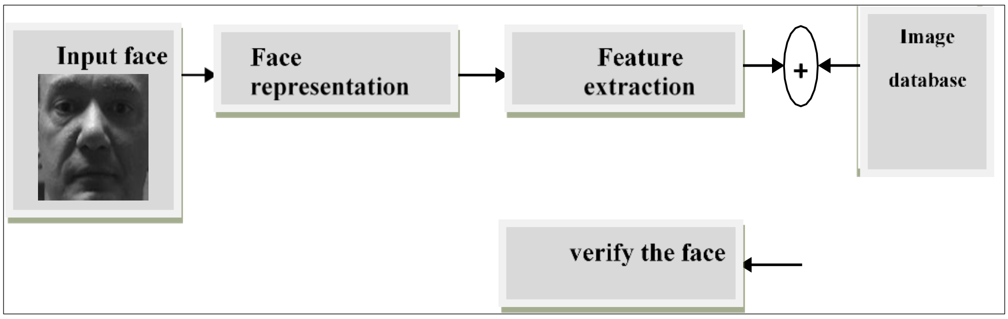


**Fig 3:** Principle of a verification process with face recognition

**Deep Neural Network Feature Extraction**

Deep Learning is type of machine learning algorithms, which use a multi-layer design to gradually extract higher level features from raw input, viz., feed-forward networks. The layers consist of neurons which translate the input signal from the previous layer to produce output signal in the same manner independently, either linear or non-linear. Any neuron from different layers can be connected to each neuron forming a network. An architecture comprising the units called neurons is a neural network. These architectures typically consist of three different layers as follows: the input layer containing the input feature vector; the output layer contains the neural network response; the input layer which is a neural network response; and the hidden layer in between is related to the input and output. Artificial neural networks consist of three fundamental characteristics: network architecture, input, functions for activation and input weight between connections. At the outset of the training stage, the network architecture and functions are selected. The output of the neural network depends on the weight value. During the training phase, the weights are tuned to produce the desired performance. In this paper, the Alexie architecture has been used [24] which consists of 5 coevolutionary layers some with max pooling layers and three fully-connected layers which have a final 1000-way SoftMax layer. It has 60 million parameters and 650,000 neurons. A 1000 way soft max, which divides over 1000 class labels, is provided with the output of the last fully connected layer. It uses nonsaturating neurons and very efficient GPU implementation to speed up training. A dropout method is used to reduce the over-fitting in the fully-connected layers. The architecture of the network contains eight learned layers five convolutional and three fully-connected which have 4096 neurons each.

**Face Datasets**

The complex face recognition system had been exposed to an array of differing facial expressions utilizing two alternative strategies demonstrated above. Countenances held within databases comparable to Face 94, Face 95, Face 96 and visage grimaces were employed for preparation and testing of convolutional neural network constructions as well as FRGC was leveraged via a traditional handcrafted element extraction approach. The Face 94 and Face 96 included 20 trials of each subject totaling 151 subjects. Face 95 was composed of 20 trials for every subject encompassing 72 people. Visage grimaces contained 20 trials of each representing 18 individuals. Furthermore, the FRGC database included 50,000 imagery of faces originating from approximately 4000 subjects. The neural system examined each encounter with fluctuating levels of intricacy, at some point analyzing subtle facial motions while at other times concentrating on more evident transformations of the lips and eyebrows. This two-pronged technique empowered both profound and broad perception of the diverse communicative signs conveyed by the human face.

**Experimentation and Results**

**Handcrafted Feature Extraction Results**

The experiment was conducted using images from the FRGC dataset, where 1480 facial photos belonging to 370 unique individuals were utilized. Each individual contributed 4 images of their face. For this study, one photo served as the template or identity for each person, while the other 3 images acted as tests to match against the template. Additionally, all images were converted to grayscale. A cropped 100x100 pixel square containing the face was extracted from each full image. Next, each of the preprocessing techniques (HE, MSR, DOG) were applied individually to the entire dataset. Local Binary Patterns analysis was then performed on all 1480 facial photos to derive a facial feature matrix representation. Matching scores between representation pairs helped determine relative similarity. In this instance, theLocal Binary Patterns feature matrices for the 370 individual templates were compared against the 1110 test images using K-Nearest Neighbors classification to generate distancescore matches between each template and corresponding tests. This paper uses Euclidean distance to compute the template-test matching scores (distances). With the obtained scoring matrices, the ability to improve preprocessing algorithms by combining score outputs was evaluated. The average score across the three preprocessing methods was considered.

The results obtained were intriguing. Here, three types of filters were utilized as a data massaging technique: the high-emphasis, mean-shifted reductors, and derivative-of-gaussian filters. The findings illuminated that it consumes 437.049869 seconds to preliminarily prepare the photos within the archives. Subsequently, local binary patterns and k-nearest neighbors were leveraged for characteristic extraction and matching to authenticate the individual. As Figure 5 illustrates, contrasting preprocessing algorithms and their fusion against the collective standard score, the average of the 3 preliminary handlers outperforms at least two of them (derivative of gaussian and mean-shifted reductors) and demonstrates competitive execution opposite the high-emphasis preprocessing method. Yet simultaneously, certain longer term images required significantly more preprocessing time and exposed unexpected correlations between filter outputs.

The receiver operating characteristic curve represents how true acceptance rate interacts with false acceptance rate through plotting verification successes versus associated improper allowances. False acceptance rate quantifies the probability of an unauthorized individual being incorrectly granted access. True acceptance rate measures the likelihood of proper access being afforded to an authorized person. Generally, the receiver operating characteristic curve visualizes this balanced relationship. Ideally, with zero percent false acceptance, biometric authentication would still authenticate legitimately one hundred percent of the time. Unfortunately, perfection is implausible, thus administrators must compromise by establishing an ideal system threshold according to usage. Determining the cutoff proves tricky since verification and false acceptance rates independently fluctuate. If the face reader's limit heightened, authenticated attempts dwindled as did false allowances. But lowering the limit boosted verified accessions while improperly approved requests simultaneously swelled. The operating point traverses its receiver operating characteristic curve through adjusting the system cutoff. Consequently, the curve charts true positive rate opposite false positive rate regarding potential test thresholds.

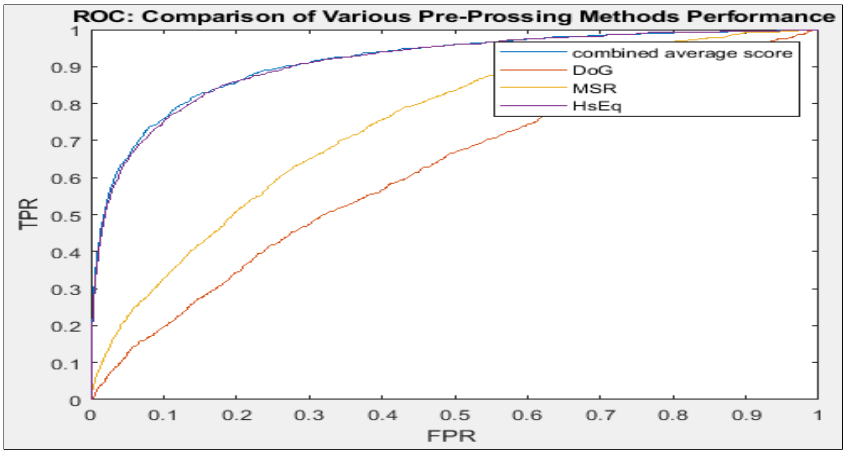


**Fig 4:** A comparison of the performance of different pre-processing algorithms and their fusion versus the combined average score.

**Convolutional Neural Network Feature Extraction Results**

The ALEXNET architecture has been traditionally utilized as a pre-trained model to assess various datasets including FACE 94, FACE 95, and FACE 96 as well as facial expressions. A mini-batch size of 10 and maximum epochs of 5 were chosen as hyperparameters in addition to Stochastic Gradient Descent with Momentum as the optimizer. To gradually absorb changes in pre-existing layers throughout training while allowing novel layers to learn more rapidly, the starting learning rate was set to a rate of 10-4 whereas the weight learning factor for new layers was increased to 20 times that of transferred layers.

In the current work, 70% of samples from each subject have been dedicated to model training with the remaining 30% held out for testing purposes. It bears mentioning that the dataset was initially separated into training and testing sets before implementing the listed augmentation techniques to ensure no overlap between subsets since the data was partitioned at the outset. Images are automatically scaled to the input dimensions required by each pre-trained network prior to both the training and testing phases. Nonetheless, convolutional neural networks as a form of deep learning necessitate labeled data to facilitate the training process of the proposed framework.

**Testing CNN Using Face 94 Dataset**

The deep network processed an immense ocean of imagery, plumbing neural depths on 94 datasets consisting solely of male facial portraits. As Figure 6 indicates, a perfectly validated 100% accuracy emerged after 28 minutes and 44 seconds of immersive training. Clearly more testing was merited to gauge how precisely the model might perform when faced with novel instances. Figure 7 therefore features a quartet of evaluation images extract from the 94 collection, along with predicted labels exhibiting the network's flawless classification of the test set. In breadth and perceptiveness, the model proved itself most adept.

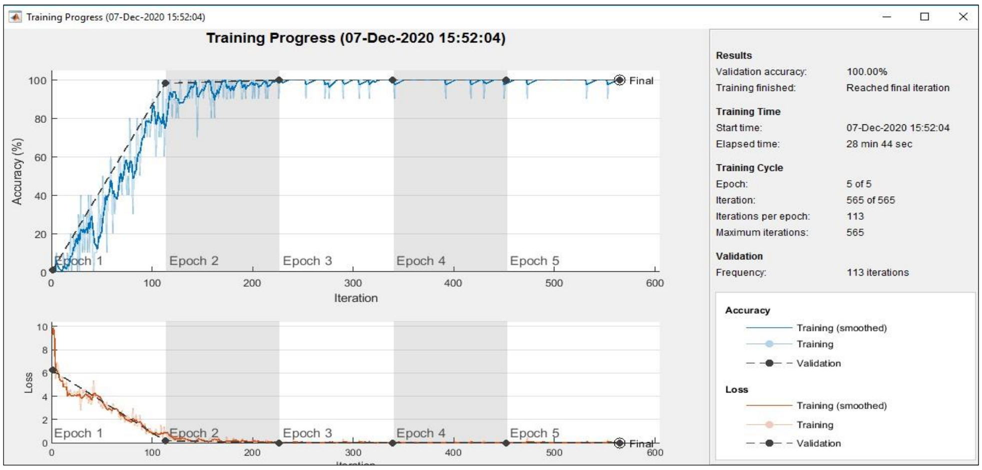


**Fig 5:** CNN training run over 94 dataset.

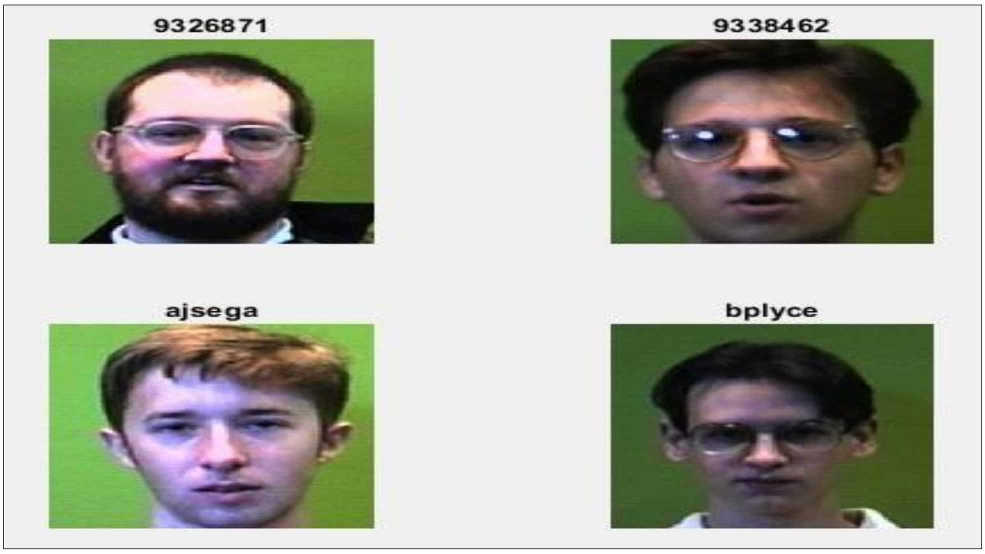


**Fig 6:** Four sample test images with their predicted labels

Understanding how convolutional neural networks process information requires examining intermediate activations at each layer. Feature maps output by convolution, relu, and pooling operations depict how successive transforms input. Crucial filters reveal how inputs decompose into independent patterns. Each channel encodes relatively features, so visualizing maps plots every channel separately. The first convolutional layer on FACE94 databases, Figure 8 shows activations processes and the first convolutional layer activation on a specific channel is shown in Figure 9. Clearly, full shapes remain visible, though several filters activate little. At this early stage, activations retain almost all initial image information. Deeper layers produce increasingly abstract, less interpretable activations, as Figures 10 and 11 demonstrate. Progressive transformation obscures original data.

**Fig 7:** First convolutional layer activation on Face 94 data base

**Fig 8**: First convolutional layer activation on specific channel (channel 10).

**Fig 9**: First RELU activation layer

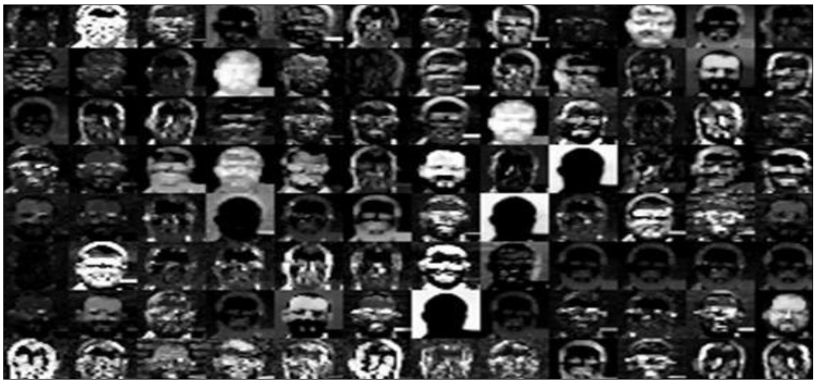

**Fig 10**

**Testing CNN Using Face95 Dataset.**

This section focuses on CNN training above statement about the burglaries on pawnshops on the morning of the new year run over Face95 databases. From the below figure, A validation accuracy of 98.61% is reached with a training time of. To see how well the network performs, an additional test set has been used. The below figure shows four sample test images from Face95dataset with their predicted labels. The test accuracy has reached at 98.38%.

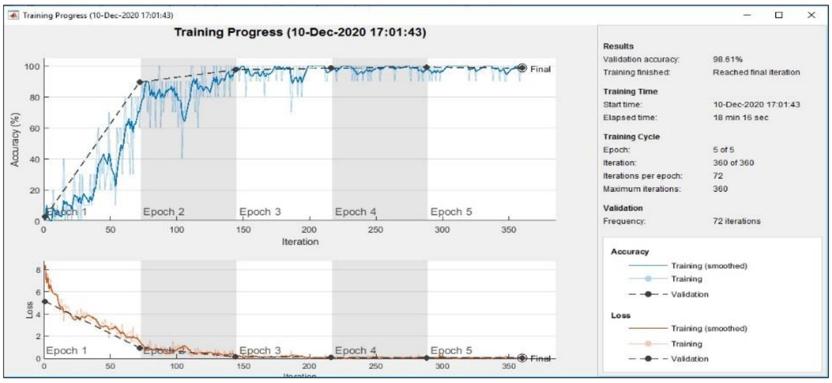


**Fig 11:** CNN training run over Face95 database.


**Fig 12:** Four sample of the test images with their predicted label.

**Testing CNN Using Face 96 Dataset**

This section conducts CNN training on 96 Face databases. According to Figure 14, it is quite evident that a validation accuracy of 99.49% was significantly achieved. The elapsed training time was 32min 16sec. In addition, to check on how the built network performed, an extra test set was used. Figure 15 shows four sample test images from Face96 dataset and their predicted label. A test accuracy of 98.54% was ultimately achieved.

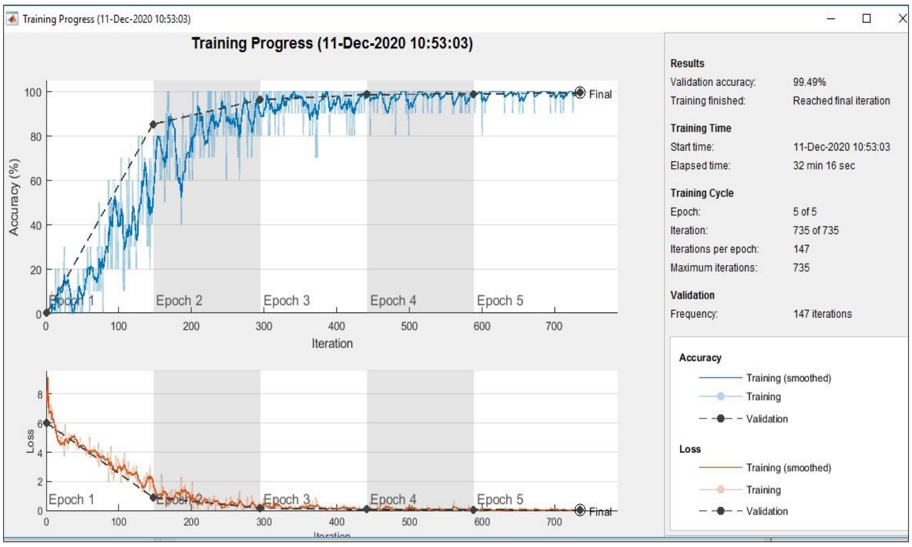


**Fig 13:** CNN training run over Face96 dataset.


**Fig 14:** Shows an four sample test images with their predicted label.

**Testing CNN Using The Face Grimace Dataset.**
The training of CNN was carried out on the grimace face database. It is observed from Fig. 16 that the validation accuracy reaches A 100% with training time of 5min 2sec. A test set can be used to see how well the network performs. The four examples of the test images are shown in the Fig. 17 with their predicted labels. It is observed that the labeling resulted in the test accuracy of 100%.

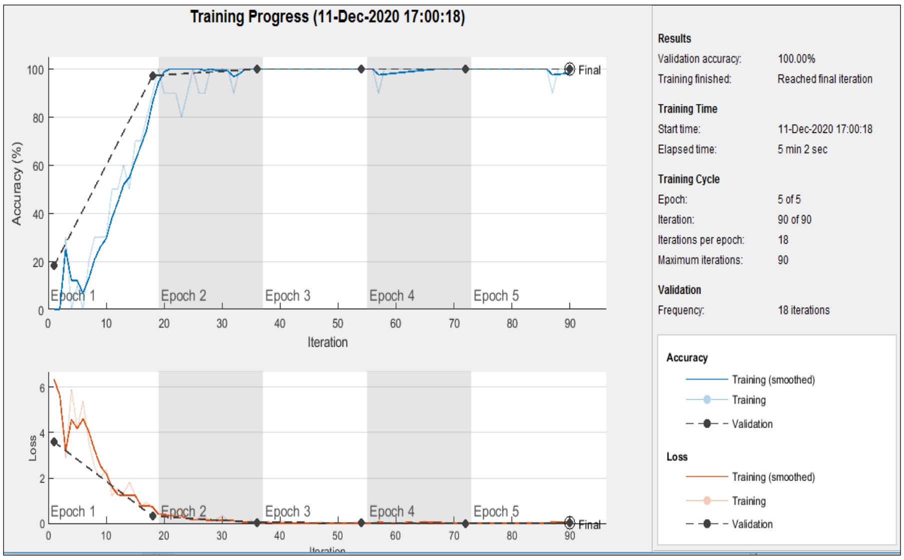


**Fig 15:** CNN training run over Face grimace dataset.


**Fig 16:** Shows an four sample test images with their predicted labels.

The training of CNN was carried out on the grimace face database. It is observed from Fig. 16 that the validation accuracy reaches A 100% with training time of 5min 2sec. A test set can be used to see how well the network performs. The four examples of the test images are shown in the Fig. 17 with their predicted labels. It is observed that the labeling resulted in the test accuracy of 100%.

**Results Discussion**

It seems to be obvious that DNN analysis method is better than the handcraft analysis method. As a result, in the handcrafted feature extraction method, we utilized three pre-processing filters to enhance the extraction performance. Moreover, and we take the average score of these three pre-processing filters where is the idea of combining the pre-processing algorithms may show a great potential. Therefore, it is clear that the average score of the 3 preprocessing algorithms is better than at least 2 of them and show a competitive performance against the HE pre-processing algorithm, this appears in the Figure 5.

Indeed, the automated section describe that we used the proposed face recognition using convolutional neural networks approach over the selected Face94, Face95, Face96, Face Grimace databases. Thus, the presented overall classification model is evaluated by common measures such as Sensitivity, Accuracy, Precision, Recall, F-Measure, and G-mean. Hence, the following Table 2 displays the set of evaluation metrics for the overall performance of the CNN proposed model. It is evident and as expected that Face94 and Face grimace that has the highest accuracy 100%. However, Face 95 and Face 96 achieves a high accuracy is about 98.38%, 98.54% respectively. Finally, all these databases have the same Sensitivity WHICH is described in the paper as ‘measures the proportion of actual positive cases that got predicted as positive.

**Table 2**: Set of evaluation metrics for overall performance of the CNN model.

| | **Face 94** | **Face 95** | **Face 96** | **Face Grimace** |
|---|---|---|---|---|
| Accuracy | 1.0000 | 0.9838 | 0.9854 | 1.0000 |
| Sensitivity | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Precision | 0.2222 | 0.2222 | 0.2322 | 0.2422 |
| Recall | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| F1-Measure | 0.3636 | 0.3365 | 0.3801 | 0.3433 |
| G-Mean | 0.9899 | 0.9750 | 0.9870 | 0.9578 |

The results of CNN performance comparison are shown in Table 3 based on training and test accuracy achievements for different Face94, Face95, face96, face grimace databases with different optimizers. It is observed that the accuracy is more compared to ADAM optimizer for all databases with the SGDM optimizer. The best accuracy in this study was 100% that occurred through SGDM optimizer for face94 and face grimace while ADAM optimizer has showed this accuracy of 100% only on face grimace. As shown in Table 3, the test accuracy of larger and Face95 and Face96 is greater by 0.44% and 4.86%, 4.73% in the case of face94 with the use of SGDM. The two SGDM and ADAM have been shown to have similar performance for the face grimace.

**Table 3:** Training & Test Accuracy performance of CNN for different optimizers

| | **ADAM optimizer** | | **SGDM optimizer** | |
|---|---|---|---|---|
| | Training accuracy | Test accuracy | Training accuracy | Test accuracy |
| Face 94 | 99.78% | 99.56% | 100% | 100% |
| Face 95 | 94.44% | 93.52% | 98.61% | 98.38% |
| Face 96 | 92.23% | 93.81% | 99.49% | 98.54% |
| Face Grimace | 100% | 100% | 100% | 100% |

The elapsed time of the CNN training has been compared between different optimizers and the same CPU computational option as shown in Table 4. The SGDM optimizer is better than ADAM optimizer, where, the training using the SGDM optimizer take less time in all datasets.

**Table 4:** Elapsed time of CNN training for different optimizers.

| | **ADAM optimizer** | **SGDM optimizer** |
|---|---|---|
| Face 94 | 51min10sec 00:51:10 | 26min44sec 00:26:44 |
| Face 95 | 19min47sec 00:19:47 | 18min15sec 00:18:15 |
| Face 96 | 43min42sec 00:43:42 | 32min16sec 00:32:16 |
| Face Grimace | 5min54sec 00:05:04 | 5min2sec 00:05:02 |

The elapsed time of the CNN training has been compared between different optimizers and the same CPU computational option as shown in Table 4. The SGDM optimizer is better than ADAM optimizer, where, the training using the SGDM optimizer take less time in all datasets.

**Conclusion**

In this study, we started with studying face recognition, feature extraction techniques, handcrafted feature extraction technique, and the deep neural network for feature extraction. Based on these methods, our face recognition model has been developed. Face recognition Models over two different models have been studied. By using these methods and other more databases with different expressions, the face recognition performance has been evaluated. It was shown how the Performance is changing with respect to the feature extraction method and the databases that it used. The parameters controlling the performance is studied and explained. Finally, the ROC curve for the handcrafted technique has been calculated. also, ROC curve, Precision, accuracy, sensitivity, recall, F1-measure and G-Mean have been calculated in the DNN feature extraction technique with different databases. Moreover, In this thesis, we investigate the performance of the facial image preprocessing algorithms of DoG, MSR, and HE. Besides we used the KNN to classify humans and LBP to extract facial automated features. In the handcrafted section of experiments, the average score of the 3 preprocessing algorithms is better than at least two of them (DoG and MSR) and shows a competitive performance against the HE pre-processing algorithm. The idea of combining pre-processing algorithms sounds promising. We continue to consider other fusion methods to deeply investigate the usefulness of the fusion. Hence the histogram equalization showed a good performance, it is useful to fuse other algorithms with the histogram equalization in pairs either considering the average or weighted combination in a linear form. The fusion could be useful not only on the matching score level but also in the pre-processing stage because each algorithm provides an advantage. Meanwhile using the raw facial databases being automatically processed proved the reliability and robustness of using CNN models over the classical handcrafted methods.

**Declaration of Competing Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Funding Information**

"No funding was received from any financial organization to conduct this research."

**References**

1. Agrawal A, Singh YN. An efficient approach for face recognition in uncontrolled environment. Multimed Tools Appl. 2017;76(3):3751-60.
2. Agrawal AK, Singh YN. Unconstrained face recognition using deep convolution neural network. Int J Inf Comput Secur. 2020;12(2-3):332-48.
3. Ahmed MAO, Hasan MM, Islam MS, Abdullah-Al-Nahid. Multi-filter score-level fusion for fingerprint verification. In: International Conference on Advanced Machine Learning Technologies and Applications. Cham: Springer; 2018. p. 624-33.
4. Ahonen T, Hadid A, Pietikainen M. Face description with local binary patterns: Application to face recognition. IEEE Trans Pattern Anal Mach Intell. 2006;28(12):2037-41.
5. Ahonen T, Hadid A, Pietikainen M, Zhao G. Face recognition based on the appearance of local regions. In: Proc 17th Int Conf Pattern Recognition. Vol. 1. IEEE; 2004. p. 153-6.
6. Alpaslan N, Hanbay K. Multi-resolution intrinsic texture geometry-based local binary pattern for texture classification. IEEE Access. 2020;8:54415-30.
7. Al-Osaimi FR, Bennamoun M, Mian A. Spatially optimized data-level fusion of texture and shape for face recognition. IEEE Trans Image Process. 2012;21(3):859-72.
8. Birch P, Young R, Chatwin C, Farsari M, Budgett D. Approximate bandpass and frequency response models of the difference of Gaussian filter. Opt Commun. 2010;283(24):4942-8.
9. Bu X, Wu Y, Gao Z, Jia Y. Deep convolutional network with locality and sparsity constraints for texture classification. Pattern Recognit. 2019;91:34-46.
10. Chaki J, Dey N. Texture feature extraction techniques for image recognition. Singapore: Springer; 2020.
11. Chen D, Cao X, Wen F, Sun J. Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In: Proc IEEE Conf Comput Vis Pattern Recognit. 2013. p. 3025-32.
12. Crosswhite N, Byrnes J, Stauffer C, Parkhi OM, Cao Q, Zisserman A. Template adaptation for face verification and identification. Image Vis Comput. 2018;79:35-48.
13. Ding C, Tao D. A comprehensive survey on pose-invariant face recognition. ACM Trans Intell Syst Technol. 2016;7(3):1-42.
14. Gonzalez RC, Woods RE, Masters BR. Digital image processing. 3rd ed. Upper Saddle River (NJ): Pearson Prentice Hall; 2008.
15. Gottumukkal R, Asari VK. An improved face recognition technique based on modular PCA approach. Pattern Recognit Lett. 2004;25(4):429-36.
16. Huang C, Zhu S, Yu K. Large scale strongly supervised ensemble metric learning, with applications to face verification and retrieval. arXiv. 2012;arXiv:1212.6094.
17. Jobson DJ, Rahman Z, Woodell GA. A multiscale retinex for bridging the gap between color images and the human observation of scenes. IEEE Trans Image Process. 1997;6(7):965-76.
18. Kaemarungsi K, Krishnamurthy P. Modeling of indoor positioning systems based on location fingerprinting. In: IEEE INFOCOM 2004. IEEE; 2004. p. 1012-22.
19. Kashyap RL, Khotanzad A. A model-based method for rotation invariant texture classification. IEEE Trans Pattern Anal Mach Intell. 1986;8(4):472-81.
20. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. 2012;25:1097-105.
21. Kumar R, Sharma HS. Comparative study of CLAHE DSIHE & DHE schemes. Int J Res Manag Sci Technol. 2008;1(1).
22. Le TH. Applying artificial neural networks for face recognition. Adv Artif Neural Syst. 2011;2011:673016.
23. Learned-Miller E, Huang GB, RoyChowdhury A, Li H, Hua G. Labeled faces in the wild: A survey. In: Advances in Face Detection and Facial Image Analysis. Cham: Springer; 2016. p. 189-248.
24. Li Z, Park U, Jain AK. A discriminative model for age invariant face recognition. IEEE Trans Inf Forensics Secur. 2011;6(3):1028-37.
25. Liu DH, Lam KM, Shen LS. Illumination invariant face recognition. Pattern Recognit. 2005;38(10):1705-16.
26. Lu C, Tang X. Surpassing human-level face verification performance on LFW with GaussianFace. In: Proc AAAI Conf Artif Intell. 2015. p. 3811-9.
27. Misra S, Li H, He J. Machine learning for subsurface characterization. Cambridge (MA): Gulf Professional Publishing; 2019.
28. Moraru L, Moldovanu S, Dimitrievici LT, Dey N, Ashour AS, Shi F, *et al*. Gaussian mixture model for texture characterization with application to brain DTI images. J Adv Res. 2019;16:15-23.
29. Nanni L, Ghidoni S, Brahnam S. Handcrafted vs. non-handcrafted features for computer vision classification.

Pattern Recognit. 2017;71:158-72.
30. Ohanian PP, Dubes RC. Performance evaluation for four classes of textural features. Pattern Recognit. 1992;25(8):819-33.
31. Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured distributions. Pattern Recognit. 1996;29(1):51-59.
32. Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell. 2002;24(7):971-87.
33. Park U, Tong Y, Jain AK. Age-invariant face recognition. IEEE Trans Pattern Anal Mach Intell. 2010;32(5):947-54.
34. Parkhi OM, Vedaldi A, Zisserman A. Deep face recognition. Proc Br Mach Vis Conf. 2015:41.1-41.12.
35. Petro AB, Sbert C, Morel JM. Multiscale retinex. Image Process On Line. 2014;4:71-88.
36. Phillips PJ, Flynn PJ, Scruggs T, Bowyer KW, Chang J, Hoffman K, *et al*. Overview of the face recognition grand challenge. In: Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. IEEE; 2005. p. 947-54.
37. Rahim MA, Hossain MN, Wahid T, Azam MS. Face recognition using local binary patterns (LBP). Glob J Comput Sci Technol. 2013;13(4):9-13.
38. Rane ME, Bhadade US. Multimodal score level fusion for recognition using face and palmprint. Int J Electr Eng Educ. 2020;57(4):344-63.
39. Romeny BMH, editor. Geometry-driven diffusion in computer vision. Vol. 1. Berlin: Springer; 2013.
40. Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering. In: Proc IEEE Conf Comput Vis Pattern Recognit. 2015. p. 815-23.
41. Speis A, Healey G. Feature extraction for texture discrimination via random field models with random spatial interaction. IEEE Trans Image Process. 1996;5(4):635-45.
42. Subudhi P, Mukhopadhyay S. An efficient graph reduction framework for interactive texture segmentation. Signal Process Image Commun. 2019;74:42-53.
43. Sun Y, Chen Y, Wang X, Tang X. Deep learning face representation by joint identification-verification. Adv Neural Inf Process Syst. 2014;27:1988-96.
44. Sun Y, Liang D, Wang X, Tang X. DeepID3: Face recognition with very deep neural networks. arXiv. 2015;arXiv:1502.00873.
45. Sun Y, Wang X, Tang X. Deeply learned face representations are sparse, selective, and robust. In: Proc IEEE Conf Comput Vis Pattern Recognit. 2015. p. 2892-900.
46. Taigman Y, Yang M, Ranzato M, Wolf L. DeepFace: Closing the gap to human-level performance in face verification. In: Proc IEEE Conf Comput Vis Pattern Recognit. 2014. p. 1701-8.
47. Taigman Y, Yang M, Ranzato M, Wolf L. Web-scale training for face identification. In: Proc IEEE Conf Comput Vis Pattern Recognit. 2015. p. 2746-54.
48. Tan X, Triggs B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans Image Process. 2010;19(6):1635-50.
49. Trigueros DS, Meng L, Hartnett M. Face recognition: From traditional to deep learning methods. arXiv. 2018;arXiv:1811.00116.
50. Varma M, Zisserman A. A statistical approach to texture classification from single images. Int J Comput Vis. 2005;62(1-2):61-81.
51. Wang Y, Xu C, Cheng J, Gao L, Sang N, Hao Y. Morphological segmentation analysis and texture-based support vector machines classification on mice liver fibrosis microscopic images. Curr Bioinform. 2019;14(4):282-94.
52. Whitaker RT. Characterizing first and second-order patches using geometry-limited diffusion. In: Biennial International Conference on Information Processing in Medical Imaging. Berlin: Springer; 1993. p. 149-67.
53. Wolf L, Hassner T, Maoz I. Face recognition in unconstrained videos with matched background similarity. In: CVPR 2011. IEEE; 2011. p. 529-34.
54. Xie S, Shan S, Chen X, Chen J. Learned local Gabor patterns for face representation and recognition. Signal Process. 2009;89(12):2333-44.
55. Zhao W, Chellappa R. Robust face recognition using symmetric shape-from-shading. Technical Report. College Park (MD): Computer Vision Laboratory, Center for Automation Research, University of Maryland; 1999.

**How to Cite This Article**

AL-Afeef LA, Al-Otum HM. A comparative study of face recognition and detection mechanisms through deep and machine learning and handcrafted features. Int J Future Eng Innov. 2026;3(4):52-64. doi:10.54660/IJFEI.2026.3.4.52-64.