REPUBLIC OF TÜRKİYE

ALTINBAŞ UNIVERSITY

Institute of Graduate Studies

Electrical and Computer Engineering

AN INVESTIGATIONAL FW-MPM-LSTM APPROACH FOR FACE RECOGNITION USING DEFECTIVE DATA

Baraa Adil MAHMOOD

Doctor of Philosophy

Supervisor

Asst. Prof. Dr. Sefer KURNAZ

Istanbul, 2023

AN INVESTIGATIONAL FW-MPM-LSTM APPROACH FOR FACE RECOGNITION USING DEFECTIVE DATA

Baraa Adil MAHMOOD

Electrical and Computer Engineering

Doctor of Philosophy

ALTINBAŞ UNIVERSITY

2023

The thesis titled AN INVESTIGATIONAL FW-MPM-LSTM APPROACH FOR FACE RECOGNITION USING DEFECTIVE DATA prepared by BARAA ADIL MAHMOOD and submitted on 17.08.2022 has **been accepted annomisly** for the degree of Science PhD in Electrical and Computer engineering.

Asst. Prof. Dr. Sefer KURNAZ

Supervisor

Thesis Defense Committee Members:

Asst. Prof. Dr. Sefer KURNAZ	Department of Computer Engineering, Altinbaş University	______________
Prof. Dr. Osman Nuri UÇAN	Department of Electrical Electronic Engineering, Altinbaş University	______________
Asst. Prof. Dr. Abdullahi Abdu IBRAHIM	Department of Computer Engineering, Altinbaş University	______________
Asst. Prof. Dr. Serdar KARGIN	Department of Biomedical Engineering, Arel University	______________
Asst. Prof. Dr. Zeynep ALTAN	Department of Software Engineering, Beykent University	______________

I hereby declare that this dissertation meets all format and submission requirements for a Ph.D. dissertation.

Submission date of the thesis to Institute of Graduate Studies: __/__/___

I hereby declare that all information/data presented in this graduation project has been obtained in full accordance with academic rules and ethical conduct. I also declare all unoriginal materials and conclusions have been cited in the text and all references mentioned in the Reference List have been cited in the text, and vice versa as required by the abovementioned rules and conduct.

Baraa MAHMOOD

Signature

DEDICATION

I sincerely thank my great supervisor Dr. Sefer KURNAZ for helping me conduct the research and prepare the thesis. This thesis is wholeheartedly dedicated to my beloved family, who has been my source of inspiration.

ABSTRACT

AN INVESTIGATIONAL FW-MPM-LSTM APPROACH FOR FACE RECOGNITION USING DEFECTIVE DATA

MAHMOOD, Baraa

Ph.D., Electrical and Computer Engineering, Altınbaş University,

Supervisor: Asst. Prof. Dr. Sefer KURNAZ

Date: April / 2023

Pages: 60

Facial recognition systems are listed as a biometric system, because they are directly related to the facial features and characteristics. They are also based on the principles of image processing, machine vision and sometimes machine learning. Face recognition systems may consider imperfect information from images. In this case, it is essential to provide a series of image reconstruction mechanisms for matching faces. In this paper a robust method was implemented and tested on the face recognition dataset based on the image's segmentation techniques. The proposed approach is that in the pre-processing phase, image should be enhanced. The image segmentation and reconstruction step is then followed by extracting the best facial features using features such as lips, eyes, cheeks and face area. This operation is based on fractal model and wavelet transform. Next, to train and test the system, the LSTM neural network is optimized using a method called Moore Penrose Matrix which named the MPM-LSTM. The results represent the proposed approach have better performance in comparison to recent methods. The performance accuracy rate for L-SVM, L-SVM-Wo, K-SVM, K-SVM-Wo, CS, CS-Wo were obtained 98, 98.5, 95, 94.5, 98.1, and 98.3 respectively, while the proposed method is obtained as 99.58.

Keywords: Face Recognition, Imperfect Face Data, Wavelet Transform, Fractal Model, MPM-LSTM.

TABLE OF CONTENTS

Pages

ABSTRACT .... **vi**

LIST OF TABLES .... **ix**

LIST OF FIGURES .... **x**

ABBREVIATIONS .... **xi**

1. INTRODUCTION **1**

1.1 BACKGROUND 1

1.2 THESIS PROPOSAL 5

1.3 AIM OF THESIS 6

1.4 PROBLEM STATEMENT 6

1.5 THESIS CONTRIBUTION 6

1.6 DATASET 6

1.7 RESEARCH OBJECTIVE 7

1.8 ORGANIZATION OF THE THESIS 7

2. LITERATURE REVIEW **8**

2.1 BACKGROUND 8

2.2 ARTIFICIAL NEURAL NETWORK 12

2.2.1 Background 12

3. MATERIAL AND METHOD **21**

3.1 BACKGROUND 21

3.2 FACE RECOGNITION SYSTEM 22

3.3 DIVISION OF FACE RECOGNITION METHODS 22

3.3.1 Orientalist Methods 22

3.4 PRINCIPAL COMPONENT ANALYSIS (PCA) 24

3.4.1 Linear Resolution Analysis 27

3.4.2 Independent Component Analysis (ICA).. 30
3.4.3 Gabor Wavelet 31
3.5 PROPOPSED METHOD... 32
4. EXPERIMENTAL RESULTS AND DISCUSSION 37
4.1 DATA SET 37
5. CONCLUSIONS AND FUTURE WORK 45
5.1 CONCLUSIONS 45
5.2 FUTURE WORK... 46
REFERENCES 47

LIST OF TABLES

Pages

Table 2.1: A Comparison Between Some Reviewed Methods 10
Table 4.1: Evaluation Criteria of Proposed Method... 43
Table 4.2: Comparison of Proposed and Recent Methods. 43

LIST OF FIGURES

Pages

Figure 2.1: The Necessary Elements That Make up a Neuron [29] 13
Figure 2.2: A Graphical Illustration of the Computational Processes……………….……14
Figure 2.3: The Functional Activation of Neurons [34][35]. 14
Figure 2.4: Sample of an ANN That uses Feed-Forward. 15
Figure 2.5: Neural Dropout Net Model [36]. 16
Figure 2.6: An illustration of what the output of a carpooling would look like. 18
Figure 3.1. Average Face And 7 Special Images Obtained From Input Images [60]. 25
Figure 3.2: Point transfer using PCA, LDA [65]... 29
Figure 4.1: Input Image. 37
Figure 4.2: Fractal Model-Based Image Segmentation with Feature Extraction. 38
Figure 4.3: Image Reconstruction with Wavelet Transform. 39
Figure 4.4: Training and Testing with MPM-LSTM... 40
Figure 4.5: Results of Proposed Method in four Steps.. 40
Figure 4.6: Outcomes of the four-Step Recommended Technique. 41
Figure 4.7: Outcomes of the Proposed Four-Step Technique. 41
Figure 4.8: Findings from the Four-Step Recommended Method. 42
Figure 4.9: Proposed Method ROC Curve. 42
Figure 4.10: Comparison of Proposed and Recent Methods. 44

ABBREVIATIONS

ANN	:	Artificial Neural Network
GA	:	Genetic Algorithm
LBP	:	Local Binary Patter
MLBP	:	Modified Local Binary Pattern
PSO	:	Particle Swarm Optimization
PCA	:	Principal Component Analysis
SVM	:	Support Vector Machine
PSO	:	Particle Swarm Optimization
LSDL	:	Locally Sensitive Dictionary Learning
LBPH	:	Local Binary Pattern Histogram
ANN	:	Artificial Neural Network

1. INTRODUCTION

1.1 BACKGROUND

Face recognition system is a technology to identify and confirming a person from a photo. A three-dimensional face recognition system employs 3D sensors that collect data on a face's shape. A face's surface characteristics, such as the shape of the eye sockets, nose, and chin, may then be recognized using this information. The face recognition system is a system that is able to identify faces with high accuracy based on artificial intelligence technology and deep learning algorithms. Through dlib, we have useable access to a trained model. It does exactly as we need it to by returning a list of integers (face encodings) in response to the facial image we provided; comparing the face encodings of faces from other images will allow us to determine whether the inputted face corresponds to any of the faces we have images of. Bledsoe was the first person to introduce a semi-automated face recognition method in 1966. The first semi-automated method for face recognition needed the administrator to find characteristics (such as eyes, ears, nose, and mouth) on images before calculating distances and ratios to a common reference point, which were then compared to reference data. In this way, faces were categorized based on features that were marked by humans. Shortly thereafter, with the work done at Bell Labs, a vector with more than 2 features (such as mouth width, lip thickness, etc.) was developed. The features selected were largely based on assessments of the human mind and their implementation was difficult. But today computer-based face recognition is a mature and reliable mechanism that has been used extensively for many access control scenarios [1][2][3][4][5][6][7].

Face Detection is a computer program that is capable of detecting, tracking, recognition, or confirming human faces from an image or video taken using a digital camera. Although there have been many advances in the field of face recognition and facial recognition for the purpose of security, recognition and presence, there are still some issues that hinder progress to achieve or beyond human accuracy. These issues of changes in the human face, such as different lighting conditions, noise in face images, scaling, background and image background conditions, etc. Some issues impede accuracy in face recognition [8].

The human face is a complex multi-dimensional design that can pass on a lot of data about the individual, including appearance, feeling, facial highlights. Powerful and productive examination of face-related highlights is a difficult undertaking that requires time and exertion. As of late, many face acknowledgment calculations have been proposed for programmed participation the executives, for example, [9] and have been effectively carried out and applied. Likewise created calculations or some current calculations with different strategies, procedures or calculations for building face acknowledgment frameworks or applications as improved or joined in [9]. Although in the development of face recognition algorithms and systems many methods has been achieved, some of the key issues related to these algorithms / systems need to be greatly reduced or addressed to achieve human accuracy in face recognition. Face features can be invoked to realize an automated attendance management system based on reliable and accurate face recognition which can be very useful in the field of proof [10]. Face roll face attendance works by comparing the face to the enrolled face in order to track their daily attendance. Standing in front of the camera is all that is necessary to have the face instantaneously and quickly authenticated. The method works on most Android devices, so users don't need a special hardware device to use face roll. The system is designed to implement a quick and secure method of recording attendance. All allowed individuals are initially photographed by the program, which then enters the data into a database. A face coordinate structure is then used by the system to store the image. Employee hours are tracked through Attendance Management. The method is used to record the amount of time that workers work and break. It is possible to manage attendance by inserting time cards, keeping track of employee hours on paper, utilizing spreadsheets, or by using an organization's online attendance software. The main challenges of face recognition and detection systems are lighting conditions, scale, blockage, display, foreground and more. Various algorithms and methods have been proposed to address these challenges.

Biometrics seems to be becoming increasingly an integral part of personal identification technology because: The price of biometric sensors is declining and this technology is growing rapidly and the general public is aware of the power and limitations of biometrics. Different biometric systems are suitable for different applications. Therefore, choosing a biometric identification method for a particular application is a difficult task. Researchers have developed a series of characteristics for the study of biometric methods, so that the more these criteria are satisfied in a biometric method, the more desirable that method is. These criteria are:

Comprehensiveness: If we call a method as a method of biometric identification, that method emphasizes a characteristic of one person in a way that all people have this characteristic (almost "every human being has a finger").

It's far clear that no biometric identification machine is in reality the high-quality way to become aware of. interesting evaluation through the worldwide Biometric institution among unique systems primarily based on four parameters: strong point, fee, effort and time spent by way of the user during detection (effort) and consumer comfort in the course of detection (Intrusiveness). an ideal biometric machine is one that has all four of the listed parameters farthest from the center of the diagram. In another take a look at [11], by thinking about 6 biometric strategies (face, fingerprint, hand geometry, voice, eye and signature) with device-readable travel files, face functions received the best percentage of compatibility. in this examine, parameters inclusive of registration, upgrade, required hardware and public popularity are taken into consideration. Face popularity device is a biometric system that uses clever automatic techniques to understand or verify the identification of a individual based totally on physiological traits. inside the last two many years, face reputation has been the concern of huge studies into gadget imaginative and prescient and sample popularity. one of the huge applications of face reputation is inside the area of authentication and protection issues. In controlling places with large populations including airports, railway stations, subways, and so on., this method is more effective than other monitoring strategies. in this manner, unique images are taken from the faces of humans and the device have to be able to identify those humans at one-of-a-kind instances, unique gestures, exclusive directions of light emission, and so forth. distinct identification strategies require unique facts series techniques, which may be categorized according to factors inclusive of personal cooperation, accuracy, fee, and ease of size. The benefits of face recognition device may be taken into consideration in the suitable power of recognition, harmlessness, friendliness and naturalness of the method for recognizing humans. Identification with the help of face images is also a visual and non-intrusive method to identify people who do not need the cooperation of the person to use it. For this reason, it has been selected as one of the three biometric characteristics of the individual for use in the electronic passport, which has been considered by security companies and has a high commercial value in the market. Today, there are many commercial face recognition systems. On the contrary, it should be noted that it comes to machine face recognition, there are several major obstacles that reduce the success and efficiency of face recognition. One of the most common obstacles in recognizing the face is different changes in the intensity and direction of light radiation, different facial expressions, hair appearance, cosmetics, age and gender, as well as the background of the image. The systems provided for face recognition are in no way able to compete with the human visual system. Humans use both general and partial facial features for recognition. Therefore, trying to model it is a good starting point for designing successful recognition systems. The primary strategies of face recognition have been of the template matching type in that the input face (face matrix) changed into in comparison with every of the samples inside the information and the end result turned into introduced. but this method had many problems, together with high volume of records and sensitivity to facial adjustments, mild radiation and .... As a result, different strategies (such as feature extraction) emerged. In feature extraction methods, the feature vector is extracted from the face image and this vector is used instead of the image. These features can be visual, statistical or algebraic. Feature extraction based on visual properties information captures the corners and texture borders of different points in the face image. A statistical feature is about a statistical feature, a histogram is of this type. Algebraic property is algebraic information from within an image. Another type is appearance-based methods, such as machine vision models and manifold learning. The use of appearance-oriented methods to solve face recognition problems has been more prevalent among researchers than model-oriented methods. Orientalist methods use the appearance of objects to describe them. The input to these systems is usually images or two-dimensional descriptions of objects. The predominant attitude that has contributed to the growth and development of appearance methods in face recognition is the use of statistical learning methods in solving the problem of identification. For this purpose, a vector description is usually used (which is the result of putting together the brightness values of the pixels) and each face is represented by a dot in the RN space (N is the number of pixels in the image). It is clear that not all points in RN space are equivalent to faces, and face images are placed on a manifold (very complex and nonlinear) in RN. The general idea in statistical learning methods is to find a suitable mapping to reduce the dimensions of the RN space (image space) and a more appropriate description of the face manifold. This makes it possible to solve the identification problem in a smaller space. It can be argued that Manifold's view of the problem of identification is correct for many different methods of face recognition - even those that claim to mimic the identification systems of living organisms. Recently, research into face recognition has focused on displaying a face that is able to obtain relevant information that is consistent with changing brightness and facial expressions. In this dissertation, a face recognition method based on a combination of multidimensional local binary patterns and co-occurrence matrix is introduced.

1.2 THESIS PROPOSAL

This research uses image datasets with incomplete portions of the face to perform face recognition operations. The proposed approach is that there is initially a pre-processing stage and then the training and testing phase begins with a neural network called LSTM. In the pre-processing phase, image data selection is performed and then refined. Then, in the LSTM neural network by defining the input layer with a series of neurons and then the middle layer (hidden), the training and test operations are presented. In this layer, there are two other models for the operation of partial data recovery and repair and feature extraction that occur in training operations. During training, the incomplete portions of the data are recovered by wavelet transform, and in the training section, simultaneous facial extraction is performed based on a fractal model based on two principles of self-similarity and unique to dimension reduction, performs feature selection and finally features extraction. Then, during the test phase in the output layer, the appearance of the face recognition layers from incomplete data is determined and a classification is performed. The structure of the LSTM neural network will be designed to be generalizable which is after training and testing data, when other data is given for training and testing, it is capable of achieving the result that is face recognition. The dataset which will be used in this approach is the http://vis-www.cs.umass.edu/lfw link. The overall objectives of the research will be as follows:

a. Diagnosis of LSTM neural network architecture with prior VGG-Face training model.
b. Ability to repair facial images and image data with wavelet transform based detection and feature extraction based on fractal model.
c. Application of wavelet transforms and fractal model in training phase of LSTM neural network.
d. High accuracy of classification and detection of LSTM neural network images in light / gray and dark / dark spaces.
e. Increased accuracy and accuracy of face recognition compared to previous methods.

The main contribution of this paper is based on the segmentation and recognition of the face based on the MPM-LSTM method that is the robust method in the face recognition and detection.

1.3 AIM OF THESIS

The aim of this thesis is use the LSTM neural network to train and test imperfect face data, and use the Moore Penrose matrix to improve it and compare the result with the methods like MPM-LSTM. Also, show that the proposed method has high accuracy than the recent methods.

1.4 PROBLEM STATEMENT

Intelligent detection systems are nowadays used as an important issue in various parts, especially security, business, administrative, economic, military, public, etc., and cover a wide range. Therefore, proposing new solutions to cover the previous challenges is a critical issue. One of these intelligent detection systems is the face recognition from images or videos that the user can use from popular entertainment on mobile phones to super-security systems. Facial recognition systems are listed as a biometric system, because they are directly related to the facial features and characteristics. They are also based on the principles of image processing, machine vision and sometimes machine learning. Face recognition systems may consider imperfect information from images. In this case, it is essential to provide a series of image reconstruction mechanisms for matching faces.

1.5 THESIS CONTRIBUTION

The proposed approach is that in the pre-processing phase, image should be enhanced. The image segmentation and reconstruction step is then followed by extracting the best facial features using features such as lips, eyes, cheeks and face area. This operation is based on fractal model and wavelet transform. Next, to train and test the system, the LSTM neural network is optimized using a method called Moore Penrose Matrix which named the MPM-LSTM. The results represent the proposed approach have better performance in comparison to recent methods.

1.6 DATASET

The dataset which will be used in this approach is the http://vis-www.cs.umass.edu/lfw link.

1.7 RESEARCH OBJECTIVE

The overall objectives of the research will be as follows:

a. Diagnosis of LSTM neural network architecture with prior VGG-Face training model.
b. Ability to repair facial images and image data with wavelet transform based detection and feature extraction based on fractal model.
c. Application of wavelet transforms and fractal model in training phase of LSTM neural network.
d. High accuracy of classification and detection of LSTM neural network images in light / gray and dark / dark spaces.
e. Increased accuracy and accuracy of face recognition compared to previous methods.

1.8 ORGANIZATION OF THE THESIS

There are several portions to this thesis. The first segment covered the problem in detail. The facial recognition system and artificial neural network study is the subject of the second half of the research background. In the fourth section, a proposed method for feature selection as well as a suggested strategy for a face recognition system have been developed. The fourth and final sections explain the implementations and experiments, the suggested approach and how it compares to other approaches, as well as the research results and potential future directions for developing the suggested algorithm in a face recognition system.

2. LITERATURE REVIEW

In this section, the concepts related to the article for the face recognition system are reviewed. At the beginning of the section, face recognition methods is discussing then related studies in the field of face recognition system with approaches such as machine learning and deep learning are reviewed.

2.1 BACKGROUND

In [9], proposed a method called SCP for face recognition in order to solve the problem of double blindness using finite data in a training set. They used a fuzzy method of maximum integration and average integration. Their results showed that a significant margin of performance could be achieved. In [10], proposed a framework called OSPE for face recognition in different situations. For example, closed faces, face emotions, and brightness changes are some of the points used in testing them. Again, their empirical results have shown that improvements in the amount of detection are achieved by partial face data. [11], describes an approach called TPGM for improving the process of identification concerning the utilization of incomplete figures. A geometric and textural cost function was minimized by the TPGM approach. Their experiments on four face databases revealed that their strategy was now comparable than other ones. The estimation of a non-rigid transformation encoding the topological structure and the measurement of the correspondence between nodes and edges constitute the topology preserving graph matching (TPGM) approach for partial face recognition. In [12], a technique called TPGM to improve the process of recognition about the use of partial figures has been presented. The TPGM method minimized a geometric and textural cost function. The results of their experiments on four face databases showed that their approach was now better than other methods. In [13], proposed a facial change for sparse display in face recognition. On the basis of an exemplary face, they formed the basics of facial changes to separate natural and facing from different angles. Their experiments showed that significant improvements are being made to the image recognition problems of an image. In [14][15] the problem of face recognition with different lights, disguise and face is considered. They developed a new method of face recognition that extracts a dynamic subset of images and extracts distinct sections for each subject. A characterization of the differentiation components was provided by those sections in order to provide a cognitive protocol for the classification of facial images using the KNN algorithm. They applied their method to popular databases such as ORL and YaleB. The results showed that detection rates can be improved by using partial facial clues. In [16], introduced a technique called LCCR to increase discrimination of representative images. The LCCR was applied to five different databases with five remote measurements. In the case of minor faces, they covered the three main facial features, namely the right eye, nose, and mouth with the chin. The results showed that the right eyes, mouth and chin have high detection rates. In [17], showed the mechanism of perception of the human face based on facial stimuli. As of late, different strategies, methods and calculations have been joined with the conventional Local Binary Pattern (LBP) or Modified Local Binary Pattern (MLBP) to accomplish face acknowledgment and upgraded face acknowledgment precision. In [18], ongoing multi-face acknowledgment utilizing inside and out preparing on the inserted GPU framework was proposed and the technique utilized for face identification dependent on convolutional neural organization with face following and somewhere down in-face recognition calculation mode. What's more, in [19], a constant double example histogram dependent on continuous face acknowledgment was utilized to accomplish ongoing face acknowledgment tasks in low-and high-goal pictures and likewise in [20], Improved face acknowledgment utilizing Local Binary Pattern and Support Vector Machine (SVM) upgraded by the Particle Swarm Optimization (PSO) calculation. In this technique, two element extraction calculations, to be specific Principal Component Analysis (PCA) and Local Binary Pattern used to extricate highlights from the pictures. In the location cycle, the grouping of the Support Vector Machine and its advancement with the article Swarm Optimization calculation utilized. In another approach presented in [21], face recognition is performed using a modified Local Binary Pattern and Random Forest that features significance and magnification to improved facial texture classification performance and compared to traditional Local Binary Pattern. The results showed that this method was better than other methods. In [22], used optimized Local Binary Pattern along with a series of advanced image processing techniques including contrast adjustment, adaptive filters, bidirectional filters, histogram refinement and image blending to identify and detecting faces. The use of three different datasets was considered in this study and compared with other methods in terms of evaluation criteria such as accuracy. The results showed that the proposed approach was better in comparison to the previous three methods. A unique image recognition and navigation system that gives visually impaired persons with accurate and rapid signals in the form of audio so that they may travel effortlessly. ROC analysis is used to compare and examine the effectiveness of the proposed approach [22][23]. In [24][25][26], used the metaheuristic methods based on the deep learning and artificial neural network. They used the metaheuristic method for reducing the feature dimension and used in the artificial neural network and deep learning to classify the faces. Part based Sparse Representation emphatically affects characterization execution in picture acknowledgment and disposed of the issues related with non-direct conveyance of face pictures and their execution by learning word reference learning. In any case, the development of picture information areas containing seriously segregating data that is critical for grouping has not been examined by current scattered piece strategies. Also, comparable encoding results between the test tests and adjoining preparing information, limited to the portion space which didn't completely get a handle on the picture highlights with comparable picture groupings to viably catch inserted discriminant data. In [27], proposed another word reference learning approach, Kernel Locality-Sensitive Discriminative SR for face acknowledgment. In the proposed K-LSDSR, a discriminant misfortune work for gathering dependent on scattered coefficients brought into a Locally Sensitive Dictionary Learning (LSDL). Subsequent to tackling the advanced word reference, the dissipate coefficients for the examples of the picture highlight are acquired and afterward by decreasing the blunder between the first examples and again the grouping results for face acknowledgment accomplished [28]. Test results addressed that the proposed K-LSDSR altogether improves the precision of face acknowledgment contrasted with contending techniques and was strong in picture acknowledgment in different conditions. In Table 2.1, a case comparison is made between the methods studied.

Table 2.1: A Comparison Between Some Reviewed Methods

Reference	Year	Methods and Techniques	Advantages	Disadvantages
[8]	2018	Less Data Method	Improved detection accuracy	Not using standard database
[9]	2018	SCP	An optimal combination of independent variables	Fewer features are considered
[6]	2017	Cognitive Process Improvement	Enhance compatibility enhancement	No more precise criteria for evaluation have been stated
[15]	2017	Face Perception Mechanism	Facial stimuli	High runtime

Table 2.2: A Comparison Between Some Reviewed Methods "Table Continued"

[16]	2017	Partial Face Image and Features	Face recognition rate for six senses	The number of benchmarks is considered less
[13]	2016	Spars Display with Different Lightness	Extracts a dynamic subset of images	Computational complexity
[14]	2015	Increased Image Discrimination	Five different databases with five measurements	Low accuracy of image detection
[12]	2015	Sparse Display in Face Recognition	On the basis of an exemplary figure, they form the basis of facial changes	Fewer evaluation criteria are considered
[18]	2018	Local Binary Pattern (LBP) and Convolutional Neural Network (CNN)	Detect facial features accurately	Non-comparison of results based on evaluation criteria and non-mention of data type used and high computational complexity
[19]	2019	Local Binary Pattern Histogram (LBPH)	Detect facial features accurately based on facial stimuli	Non-comparison of results based on evaluation criteria and non-mention of data type used and high computational complexity
[20]	2017	Optimized Support Vector Machine (SVM) with Local Binary Pattern (LBP) and Particle Swarm Intelligence (PSO)	Accurate identification and recognition of face features and pattern extraction for face classification for high precision matching	Non-comparison of results based on evaluation criteria and non-mention of data type used and high computational complexity
[21]	2013	Modified Local Binary Pattern (M-LBP) and Random Forest	Detect facial features accurately based on facial stimuli	Non-comparison of results based on evaluation criteria and non-mention of data type

Table 2.3: A Comparison Between Some Reviewed Methods "Table Continued"

[22]	2019	Optimized Local Binary Pattern (O-LBP) based on Contrast, Adaptive Filters, Bilateral Filters and Histogram Enhancement	Using three different datasets and displaying evaluation results with previous methods and improving performance - ease of operation - high execution speed	Requires ready-made boxes - The number of attributes in each dataset is not the same, so there is no mention of the type and structure of attributes.
[27]	2020	K-LSDSR	Standard data usage - High accuracy in identifying features and facial portions of images	High Computational Complexity - Lack of Evaluation in Different Criteria

2.2 ARTIFICIAL NEURAL NETWORK

2.2.1 Background

The term "artificial neural network" (ANN) refers to a computational model that replicates the structure of the neural networks (NNs) found in the cerebrums of living organisms. These NNs are composed of a vast number of interconnected nerve cells known as neurons. Between 10 and 11 neurons make up the human brain on average. Neurons are the fundamental building blocks of all living things. The transient motivations of the signals produced with bulk of specified cell, also known as a layer, are what are responsible for establishing the correspondences between the neurons. This information is passed along from one neuron to the next in the form of electrochemical intersections, also known as neurotransmitters. The dendrites, which connect them to the cell body, are where these electrochemical crossings are situated. The impulses from several neurons are sent to nerves by the dendrites. Potentials are assessed inside the nerve cell body to decide whether or not to produce an electrical signal by the neuron. It is possible for a signal to have either an excitatory or an inhibitory effect on the neuron that it is received by after being transferred from one neuron to the next. In contrast, the excitatory effect leads the getting neuron to fire, preventing it from terminating. The inhibitory influence keeps it from doing so. No matter what, the influence of these components. Figure 2.1 shows a modified version of the basic elements of a neuron.

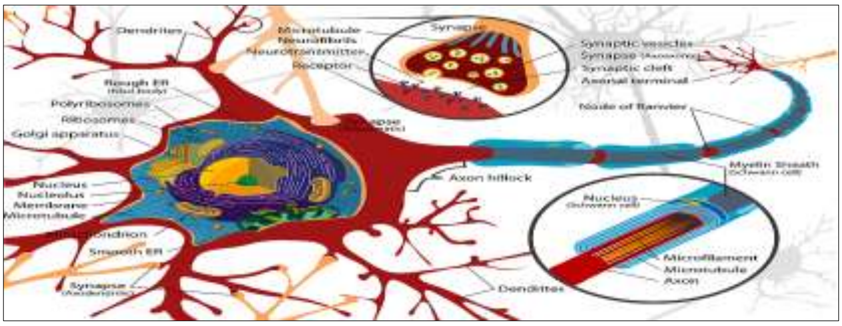

Figure 2.1: The Necessary Elements That Make up a Neuron [29]

Through the use of mathematical features known as loads, the influence of aids to the neuron is changed in ANNs. Each piece of information undergoes a process of duplication known as the comparing weight just before it is transmitted to neurons. The actuation work will use the findings of the summation to determine how the neuron will respond. This occurs when a neuron has accumulated such weighted features despite having a predisposed worth. Also, the bias values provide the neuron more freedom to alter the output value in accordance with the traits it needs to recognize. This flexibility is provided by the neuron's ability to adapt. When there is no bias applied, for example, the output of the neuron is always 0 whenever the inputs are also 0. This is the inputs that have a value of 0. As a consequence of this, the presence of bias value makes it possible for neurons to evaluate the necessary output based on the characteristic. The synopsis S of the inputs and the bias of neuron j are both displayed in Equation 2.1.

$$y = \sum_i x_i \, w_{ij} + b_j \tag{2.1}$$

xi are the contributions that are made by that neuron, wi are the loads that are placed on that neuron are the neuron's biases. Figure 2.2 provides a graphical representation in which a neuron grasps that are needed to register a request for a result.

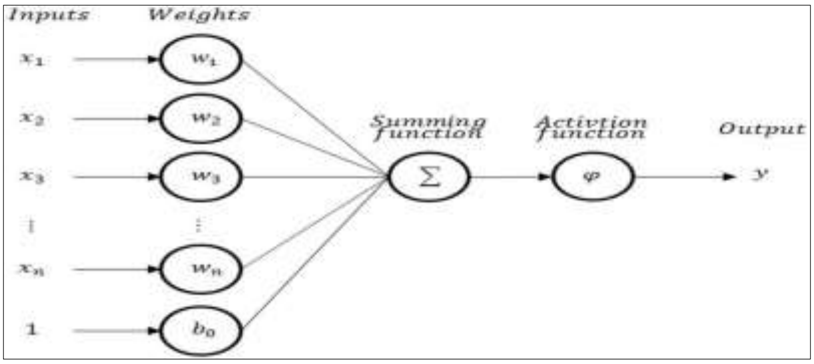

Figure 2.2: A Graphical Illustration of the Computational Processes.

The neuron is able to gain the capability of setting nonlinear boundaries for independent direction when the results of the summation are converted into an actuation effort. If an actuation job is not included the only theoretical restriction that a neuron might apply in its calculation to categorize the precise predictions. Additionally, deeper NN neurons might be able to provide more intricate restrictions for each of the classes, which could enhance NN's prediction accuracy. Additionally, the development of these intricate constraints may be aided by the use of inclination values within each neuron. By combining neuron limitations before neuron and modifying the regions of each piece of mind-boggling restriction, these inclination values are created. Three of the most prevalent initiating capacity types are as shown in Figure 2.3. (ReLU). In terms of execution, NNs that possess ReLU initiating capacities have been found to be considerably more effective than those that possess other enactment capacities. Due to the non-linear nature of the calculations, the outcome can be deduced from the contributions by identifying the necessary components and running the appropriate equations. Although hash values cannot be utilized to recover initial information [31][32][33]. Additionally, a NN could take different paths to arrive to a particular outcome.

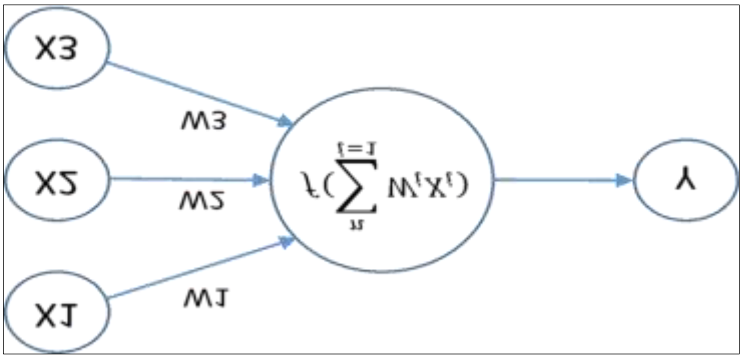

Figure 2.3: The Functional Activation of Neurons [34][35].

Layers of FFNN's neurons are dispersed across the network, and the productions of each layer's neurons are coupled to the contributions of the neurons at the layers below it. The yield layer, the information, and secret are the 3 various kinds of layers in the network. The quantity of neuron at the result layer relates to the number of results needed by the framework, and the amount of neurons in the information layer corresponds to the amount of contributions made by NN. The major layer that connected to is the information layer. This means that the maximum number of neurons that may be incorporated within such layers is constrained by the number of executions necessary to complete the job for which a NN is being used. In addition, regardless of the circumstances, completing the task at hand by utilizing only the information and result layers is impossible because doing so restricts both the number of elements and the complexity of those elements that the NN is able to identify for the purpose of producing accurate forecasts. An illustration of a totally associated feedforward deep neural architecture can be found in figure 2.4.

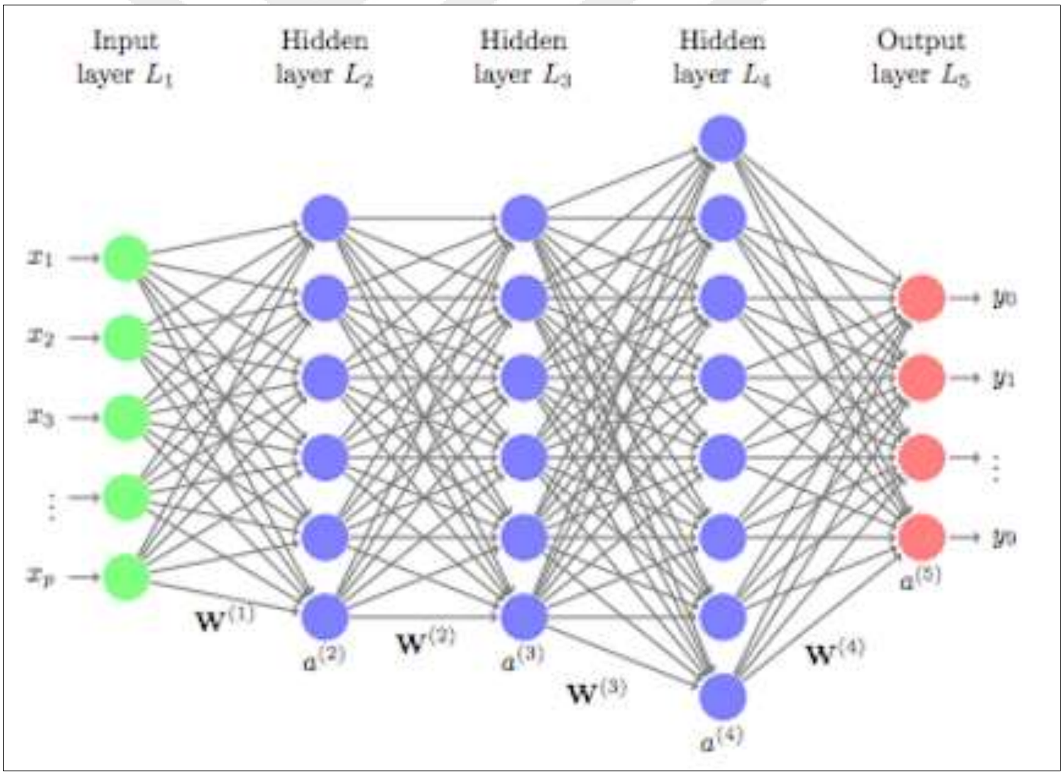

Figure 2.4: Sample of an ANN That uses Feed-Forward.

Increasing the stowed away layers makes the computations required to predict a class of information more difficult and, as a result, lengthens the amount of time needed for that prediction. The amount of mixes that activate the hidden layer's neurons will rise as the amount of layers grows, as will the amount of neurons in the hidden layer. The amount of highlights that may be seen in the buried layer will rise as a result of this. As an illustration, the intricacy of an element that may be found in deeper levels rises as the level of depth at which the network is layered increases. As a consequence of this, networks that have more than one hidden node are in a position to recognize more astounding highlights. These networks are referred to as profound NNs.

One of the primary issues that is researched using profound NNs is the phenomenon known as overfitting. Predictions in this phenomenon are dependent on explicit features in the NN, which severely restricts the predictability of such elements. As a result, there is a good chance that any further data sources that might fit that category but do not trigger the neurons connected to such items will be misclassified. It is possible to circumvent this issue by randomly removing a certain number of neurons from the hidden layer at the beginning of each cycle of the preparation stage. This compels the neural network to devise new strategies for producing the same forecast and lessens its reliance on elements that are explicitly stated. This strategy is known as Dropout, and as seen in Figure 2.5, it led to a discernible rise in the quality of the NN forecasts produced.

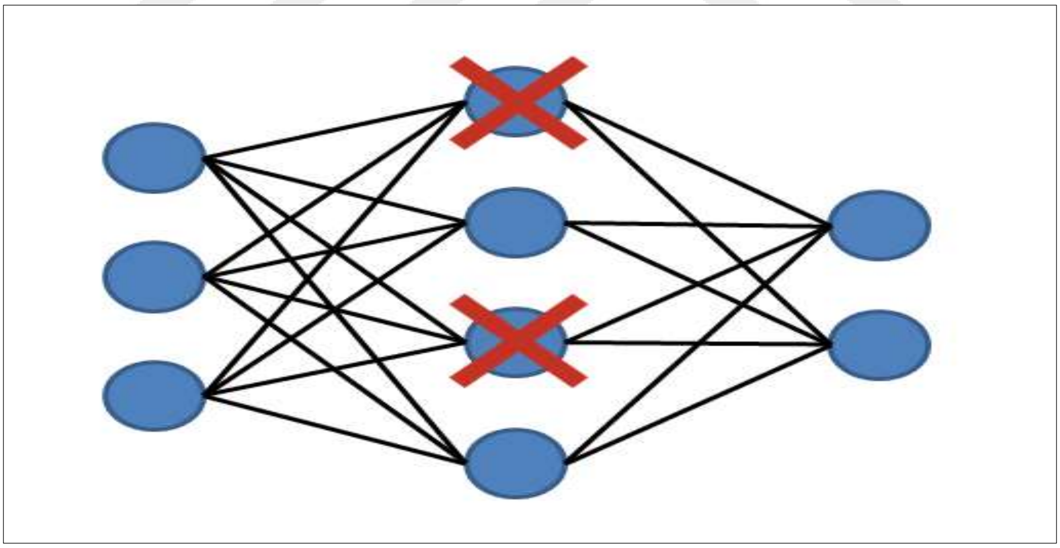

Figure 2.5: Neural Dropout Net Model [36].

The outcomes of calculations performed in each neuron, starting with the input layer and working all the way up to the result layer, are the predictions made by a NN. In addition to the information advantages of the information layers, these calculations are mostly based on the preferences and loads of the NN. The training process entails changing the inclinations and loads of the NN in order to work on increasing the accuracy of the forecasts given by such networks. Equation 2.2's Sum of Squared Errors (SSE) approach is used to estimate the expense, which is the difference between the required quality and the anticipated quality, to complete this update. An expenditure is what we're calling this upgrade.

$$C = \sum_i \frac{1}{2}(y_i - \widehat{y_i})^2 \qquad (2.2)$$

wherever y stands for the expected value and, for the actual mandatory incentive for I yields in the outcome layer [37], and where y represents the expected value [37].
It is necessary to back-spread over NN before updating the predispositions and loads' quality. This calculation takes into account the determined error in the layer beneath the fractional initial subsidiary for each weight and preference. Then the initial value of the inclination or weight is deducted from the values of the characteristics that were calculated from each subsidiary. This is true because the boundary's error-causing effect is positive. Additionally, if the value of the subsidiary is increasing NN's expectancies [38]. In earlier generations of neural networks (NNs) and those that came before them, two-layered sources of information like photographs had to be transformed into one-layered vectors so that the network could process them. The two-layered information suffers from a data shortfall as a direct consequence of this alteration because highlights in one of the aspects are made longer in proportion to the dimension of that aspect. For instance, the elements formed by vertical pixels are lost when a picture's columns are arranged closely together and its pixels are uniformly straightened. This is so because the width of the image is equal to the space between two consecutive pixels while looking up. The technique of finding highlights in two-layered data sources use NNs to get around this problem. This methodology makes it possible to locate neighborhood highlights without respect to the order in which such elements appear in the data [39].NNs are utilized for repeating activities in an inferotemporal manner in the visual arrangement of human beings. These types of networks make use of 2-layered channels which are interwoven through the entirety of the image in order to differentiate between highlights, with each channel designating a particular component for each of the layers. Because the channels themselves are composed of two layers, the recognized items are also composed of two layers. This method makes it possible to separate the highlights from the 2-layered contribution without having an effect on the state of the input [40]. As a result, channels can identify various highlights based on the information sources and the assignment requested by the NN. The main distinction between this

methodology and conventional PC vision techniques that identify particular types of elements, like corners and edges, is this. The magnitude of the steps reveals the degree of development that has occurred along each of the potential courses, and they are interlaced across the entire image. Another display is created when the load results and information values are merged with the size of the new cluster that has been decided by the established processes for layer comparison. This exhibit is created by fusing the information values and load outcomes [41].

Another essential type of layer present in NNs is the pooling layer, which is further separated into the max-pooling and typical pooling layers. Max-pooling layers, on the other hand, have demonstrated a substantial dedication to avoiding overfitting by reducing the amount of the results from one channel before sending the characteristics to the following layer. Before the results are sent, this is done. The channels are followed by the max-pooling layers, which create another cluster. The information from the channels is used to generate this cluster. It is crucial that the steps used to control how big these channels move throughout each convolution are provided in a way that is similar to the size of the channels used in these layers. An illustration of a maximum pooling layer is presented in figure 2.6. This layer has a channel size of (2x2) and steps that are also (2x2).

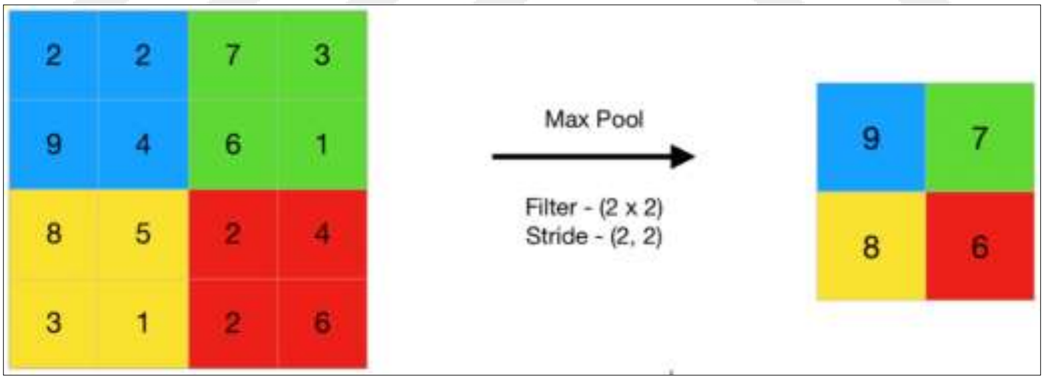

Figure 2.6: An illustration of what the output of a carpooling would look like.

The output of the last layer, which could be either a channel or a maximum pooling layer, is leveled to a single vector and fully coupled to the FFNN. This final layer is also known as the output layer. Due of the numerous channel combinations in the layers, the neurons in the fully linked layers can now detect more intricate two-layered neighborhood highlights in the input. Moreover, a three-layered contribution to NN that separates the data into exhibit tiers made up of two layers each is manageable. Each layer will then have several channel layouts created by the NN, and deeper levels will eventually be able to blend highlights from multiple layers [42]. Artificial neural networks (ANNs) rely on the dispersion of the neurons and the loads that are distributed among them to choose the best alternative, just as the geography of the biological NN are in charge of the decisions that the mind makes. To determine which choice is the best, this is done. It is possible to use two identical NNs for completely unrelated tasks, each of which would require the application of different load values across their neurons in order to arrive at a distinct set of conclusions [43].

Backpropagation is a vital component in the prevalence of NNs since it makes it possible for the network's presentation to be significantly improved. As can be seen from Equation 2.3, the process of back-propagation requires three characteristics in order to be successful in the task of refreshing loads w of an ANN. These characteristics include the pace at which the network's result advances, the rate at which bad luck is refreshed O/w, the error E between the network's result and the one that is actually needed of it, and the rate at which learning L [44].

$$\hat{w} = w - \frac{\partial O}{\partial w} \times E \times L \tag{2.3}$$

The testing inputs from the training data set are handled by the forward pass of the NN. Afterwards, the identified error esteem is utilized in the process known as backpropagation. Nevertheless, in order to regulate the delta values in lower ranges for loading refreshes, a learning rate has been implemented. This is necessary due to the fact that big blunder values might lead to huge delta values. This control for the delta values makes sure that the detonating weight values are avoided, and it also makes it possible to identify the load values that are responsible for the fundamental error [45]. The determination of the rate at which errors in the results are being made in relation to weight values might lead to the production of one of three distinct characteristics, namely: An example of a positive worth showing how being heavier makes it more difficult to judge oneself accurately. In order to desired results, the weight esteem must be decreased in this way. Using the exercise delta esteem will allow you to achieve this. A value that is less than zero, which is an indicator that the error will be reduced as a result of the increase in the value of the weight. The existing weight esteem needs to be increased by the selected delta esteem in order to bring about a reduction in the error worth and the generation of more accurate conclusions. A value of zero, indicating that there is no need to make any adjustments to the existing. In order to close the gap these potential attributes. The amount of work required to complete the ANN's assignment would be lessened as a result. However, determining the ideal load values calls for a variety of considerations, including age [46]. The ideal computation of the optimal load values, require that this be done. The preparation of the NN is dependent on a mismatch between the characteristics that the neural network provides while employing its existing predispositions and loads, and the outcomes that are necessary for this information. However, in order to make an attempt at not missing the least unfortunate event on a global scale, those borders cannot be promptly refreshed. In addition to this, because back-propagation starts from the yield layer to the information layer, adjustments to the borders. In other words, the yield layer is the layer that is updated first. Because of this, the development of a profound neural network calls for a significant investment of resources as well as the planning and execution of tests with the end goal of producing the required presentations [47][48]. The layers that are closer to the information layer are in charge of identifying the basic highlights, and these layers will probably be divided among the numerous uses that neural networks are put to. For example, if a Convolutional NN is necessary for recognizing facial highlights or characteristic mark components, channels in a layer that is closer to the information layer may be able to distinguish between comparable. When those components are blended in after the layer, the result is complex high points that are more involved with the successful completion of important tasks. In any case, during the process of preparation, the layers that are closer to the outcome layer receive more frequent refreshes. Because of this, the process of the preparation can be significantly sped up [49][50]. In the move learning process, the NN preparation does not start with nothing prepared; are not distributed randomly. The NN structure is another factor, which is distinguished by the number of the layers, the number of neurons in each of the layers, and the manner in which the layers are connected to each other, may vary from one application to the next in terms of the number of the elements that are to be recognized at each level of complexity and the level of complexity of the elements that are to be recognized in order to arrive at the expected results. After that, a neural network that has already been trained might be used for preparation, and this would be the case regardless. The preparation makes use of rough highlights that the neural network already knew for certain, as well as ensuring that pre-owned constructs are suitable for application purposes. Therefore, move learning had brought about the effect of allowing a significant decrease in the amount of time spent preparing while yet offering required presentations [51][52].

3. MATERIAL AND METHOD

3.1 BACKGROUND

Currently, the prevailing trend among the machine vision community is the use of appearance methods to solve identification problems. Face recognition as one of the main issues of identification is no exception to this trend and the volume of studies and research conducted in the development of appearance methods is much more than model methods. The reasons for this tendency, in addition to the simplicity of orthodox descriptions and their acceptable performance on many issues, also go back to our future needs. Appearance solutions do not involve many of the problems of modeling methods, for example, even model extraction for somewhat complex shapes, as well as determining the features needed to segment models into images, are difficult and unresolved issues in machine vision. Undoubtedly, the pioneer of face-oriented methods in face recognition is none other than Moschetti et al., that created a revolution in recognition technology and caused that today less than model-oriented and form-oriented methods [53]. It is less talked about in academia. It may be interesting to know that one can distinguish one object from several million objects in less than ten seconds by using visual methods [54]. However, historically the first face recognition machine (Bledsoe machine) belonged to a set of modeling methods.

Depending on the nature of the input of a face recognition system, the recognition system can also be classified into two-dimensional and three-dimensional systems. Although a simple and concise study shows that in many methods, the problem of face recognition has been introduced as the recognition of a three-dimensional object (face) from two-dimensional images, but with the growth of technology and the advent of three-dimensional face scanners, of course more attention It has become a three-dimensional information processing. Our focus in this dissertation is exclusively on methods that examine the problem of identification from two-dimensional images.

In the identification using two-dimensional images, face brightness data or color images are used. In these methods, a data camera is needed to extract the data, which can be used to capture the person's face. The advantage of this method is the ease of imaging as well as the cheapness of the data collection system. Brightness image is the most common type of data to identify people based on face image. In judicial applications, its use has the advantage that it can be understood by individuals and the observations of individuals can be used in this case. This is why face painting is used in criminal applications to identify individuals.

3.2 FACE RECOGNITION SYSTEM

A face recognition system in its simplest form, consists of two components:

a. Properties extraction section
b. Classification section

The input of the identification system is the search image C and its output is the image class (personal identity). The first detection system extracts a feature vector X = [X_1, X_2,… X_k] from the input image and compares the extracted feature vector with all the feature vectors E_1, E_2,…, E_Q learned in the training phase by a similarity criterion. Based on the results of this comparison, the exploration image class is determined. What will be discussed more in this chapter will be the feature extraction section because most of the research has targeted the feature extraction section. K - NN classifiers, SVM machines, artificial neural networks, Bayesian classifiers are just some of the machines used in face recognition research.

3.3 DIVISION OF FACE RECOGNITION METHODS

As mentioned, the methods for face recognition that have been described in previous years have been done in three ways: appearance, model and combined. Since a large amount of studies and researches in the field of face recognition in the form of appearance methods have been reported, more attention has been paid to the appearance methods.

3.3.1 Orientalist Methods

Orientalist methods in face recognition are classified into two groups: partial and generalist methods. The first category of appearance-oriented methods are partisan methods. Unlike Klinger methods, in detail methods, a feature vector is formed for each part of the face, such as the eyes, nose, and lips, and the detection machine uses the recognition results of each component in its final decision. In Klinger methods, a feature vector is extracted from the whole face image. For this purpose, a vector description (which is the result of combining the brightness values of the pixels) is generally used, and each face is represented by a point in the RN space (N is the number of pixels in the image). It is clear that not all points in RN space are equivalent to faces, and face images are placed on a manifold (very complex and nonlinear) in RN (in today's face recognition methods, each face is described by a feature vector in RN space. It is clear that not all points in the RN property space are equivalent to the image of the face;

In fact, local and general characteristics have different characteristics and are complementary in the practice of classification. Local and general properties are sensitive to various changes. For example, lighting changes have a greater effect on local features and state changes have a greater effect on overall features. Therefore, the combined methods that use both methods can increase the efficiency and generalizability of the classifiers [55]. Of course, it should be noted that local methods can be considered as a special case of combined methods. Because general information also usually participates in the algorithm.

As mentioned above, in appearance-based identification methods, images are converted to a large-dimensional vector, in which each image is represented by a point in the next n space. Due to the large size of the images, calculations become very time consuming and heavy. Therefore, in these methods, in the first step, statistical techniques are used to reduce the dimensions of the images so that they can be used to distribute the data and then, without losing basic information, the images in a sub-space with dimensions. Less, which is called the characteristic subspace, presented. In this case, with the arrival of each new image to identify it is taken to the characteristic space and the similarity of the input image with the images given in this space is checked. Thus, any image, which consists of two-dimensional data, can be expressed as an n-dimensional vector of data. To do this, they place rows or columns of data in a row. For example, a two-dimensional image with dimensions r * c can be expressed as a vector in space. This high-dimensional space can also be expressed with the help of statistical techniques under the smaller-dimensional space. The primary purpose of subspace analysis is to determine this space with smaller dimensions. Linear or non-linear subspaces can be used for this purpose. In these methods, the data is placed in a matrix after being converted to a n-dimensional vector. For example, there is an n * P matrix where each xi is an n * 1 matrix equal to the number of pixels in the image, and P represents the number of images. As mentioned, the methods are divided into two categories: general and specific. In Klinger methods, the predominant research tendency is to use statistical learning methods or in the language of face recognition, recognition in subspaces. Statistical learning methods are divided into linear and nonlinear branches. Important examples of linear methods are PCA, LDA, and ICA extensions, and important examples are the nonlinear solutions of

KPCA core machines. In the following, we will examine the prominent methods of statistical learning in face recognition. Data analysis is usually a complex task because of its large dimension. Pattern recognition systems often have difficulty analyzing high-dimensional patterns. Therefore, feature extraction before the clustering stage of patterns is done using a conversion to reduce the dimension, so that some variables are more meaningful than others and are called more meaningful components of the feature. If less important variables can be ignored, then a reduction in the desired dimension has occurred. In addition, it can be assumed that less important components among the converted variables may be due to noise. Therefore, their removal in some cases may also lead to noise removal. The feature extraction operation maps the data space to the feature space, which has a much smaller dimension than the original data space, while retaining important data of the data. The data can be reduced using methods such as linear and nonlinear principal component analysis, factor analysis, feature clustering, and so on. What is discussed here are the basic concepts of linear and nonlinear fundamental component analysis to demonstrate some of the capabilities of these two methods.

3.4 PRINCIPAL COMPONENT ANALYSIS (PCA)

PCA is a solution for mapping multidimensional data to a space with less dimension and less data loss [56][57][58][59]. The main idea of PCA is to reduce the size of the data in a data set so that changes in this data are maintained as much as possible. The basis of this method is that each image is considered as a point in a space with high dimensions. In the case of multiple images, these images are scattered dots in this high-dimensional space. Now we look for orthogonal vectors that by transferring data on those vectors, the data become as independent as possible. Therefore, this method looks for orthogonal vectors that meet this requirement as much as possible, and for this purpose, it uses principal element analysis, or PCA for short. In this method, first the data covariance matrix and then the matrix of vectors and eigenvalues are calculated. Matrices of special vectors are the orthogonal vectors that form the property subspace, and by transferring data to this subspace, the data becomes independent, which is called face space. To transfer data to this subspace, we multiply the data matrix in the matrix of special vectors or the new subspace vectors. Data transfer to this subspace has the important advantage that with this transfer, their volume is also reduced, which makes it possible to store information with a smaller volume, which is the case when the number and volume of data increases. Finds is of great importance. The special vectors used in this method for data transfer are orthogonal and are in the direction of maximum data scatter, and the expression of images in the new space is the expression of data with a minimum of squares error. For the input image in this method, the image is first transmitted using a matrix of special vectors or vectors that make up the subspace. Then, in the reduced reduction space, it is compared with the available data and the most similar image is selected as the detected image. Criteria such as Euclidean, Manhattan, Cosine, etc. can be used for comparison. The new space that is created has dimensions equal to the dimensions of the image, and if the transfer vector is resized and displayed, images similar to the face image will be obtained. In fact, the goal is to express the input image using the linear coefficient of these images. Figure 3.1 shows seven feature vectors that are equivalent to the first five eigenvalues. These images were obtained by applying the principal elements method to a portion of the FERET database images.

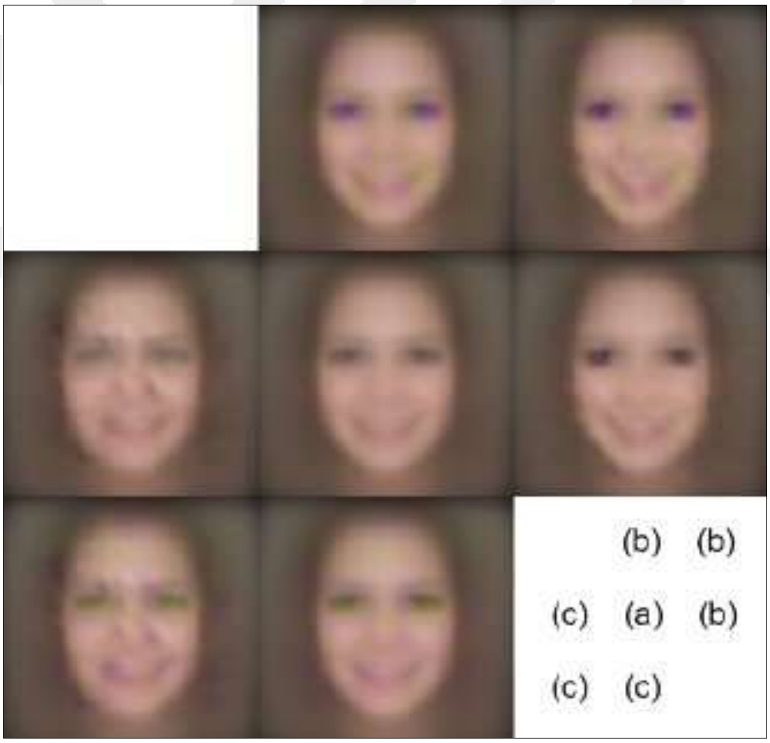

Figure 3.1. Average Face And 7 Special Images Obtained From Input Images [60].

The advantages of this method include ease of implementation and use, reduction of data volume and high speed. Ignoring the dispersion within and between classes of data and not paying attention to the image tag to identify and differentiate between different images of a person in the database, is one of the disadvantages of this method, which makes the maximum dispersion matrix, not Not only maximizes the scatter matrix between classes, which is useful for classification, but also maximizes the scatter matrix within classes, which is undesirable for classification. Also, the need to update all available information by entering a new image in the database is another disadvantage of this method. For this reason, the use of this method is more used as a dimensional method as preprocessing in other methods such as neural networks. The mathematical calculations and steps of this method are as follows:

Convert an image matrix, which is a two-dimensional matrix, to a one-dimensional vector and put it together to form a data matrix. As a result, if we have a P image with dimensions r, c, the resulting data matrix will be a matrix with dimensions m * P, where m is the total number of pixels in the image, m = r * c.

a. Calculate the average of the obtained matrix and transfer the data to center zero. To do this, after calculating the average image matrix, we subtract all images from the average matrix.
b. Calculate the covariance matrix then its vectors and eigenvalues. In this step, we obtain the covariance matrix and calculate the matrix of special vectors and its eigenvalues.
c. Transfer data matrix to new subspace using special vector matrices. To do this, we multiply the image vectors by the transfer vectors. By doing this, the original data, which are the same images, are converted into transferred data, which indicates the location of the data in the new vectors.
d. By entering a new image to identify
e. Convert image matrices to vectors and move them to center zero. To do this, we subtract the input image from the average image of the original images.
f. Transfer data vector to subspace using special vector matrices. To do this, like the operation for the original images, we multiply the new image vector by the transfer vectors.
g. Check the similarity between the transferred vector and the existing vectors and select the most similar vector. To do this, based on the defined similarity criterion, we examine the image that most closely resembles the input image. theory in algebra states that two matrices and eigenvalues are the same, and that eigenvectors are equal to eigenvectors multiplied and normalized in the matrix. With the help of this theorem, special vectors can be obtained through the matrix, which is P * P, instead of calculating through the

covariance matrix, which is N * N, because it has a lower computational load. So in step 4 we use instead of.

Continuing his research on face recognition, Pentland proposed the Eigenfeatures method for diagnosis [61]. In this method, instead of PCA expansion is applied to the whole face image, nose, eye and mouth patterns are prepared for each input image and then Carhanded-Lowe expansion is applied to these patterns. To detect, the distance between the input patterns is measured with all the patterns in the training set and a distance vector is formed. From this distance vector, one point is calculated for each person and the person who obtains the most points is introduced as the output of the system. View based Eigenface method is an extension of the classic Eigenface method to make the algorithm resistant to changing the angle of the face [62]. In this method, each person took pictures of different views (Pantland used 9 pictures in the range (_- ^ +) 90). It is stored in the training set and then a face space is formed for each angle. This way, a three-dimensional description of each face is stored in the system. In the first stage of diagnosis, the angle of the head (using mapping in different spaces of the face and calculating the minimum distance) is determined. Then, according to the angle of the head in the space of the corresponding face, the search operation is performed to find the winner. Compared to the Eigenface method, this method has a higher efficiency in detection, although the volume of calculations has also increased. Studies have shown that the Eigenface method is sensitive to changes in ambient light, image shifts and scale changes.

It is interesting to note that the method of special faces can be used to distinguish many objects. Also, traces of using PCA conversion and similar ideas with the special faces method can be seen in voice processing, remote sensing, medical engineering, character recognition, scene analysis methods, and so on.

3.5 LINEAR RESOLUTION ANALYSIS

The linear separator analysis method proposed by Belhumeur et al. [63][64] finds vectors that differentiate between classes. To do this, two dispersion matrices between class and intra-class are defined according to Equations 3.1 and 3.2.

$$S_B = \sum_{i=1}^{c} M_i . (x_i - \mu) . (x_i - \mu)^T \tag{3.1}$$

$$S_w = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_i - \mu_i).(x_i - \mu_i)^T \tag{3.2}$$

Where Mi is the number of images in class i and C is the number of classes, μ and the average of the total data and the average belong to class i and x of the individual image. Intra-class dispersion indicates the dispersion of data around the average of its class and dispersion between classes indicates the dispersion of data around the average of all classes. The goal is to increase inter-class dispersion and decrease intra-class dispersion. To do this, the property vectors of the Sw-1 * SB matrix are used.

In this method, the goal is to keep the data related to a class as close together as possible in the newly created subspace and the data of different classes as far apart as possible, in other words, the maximum dispersion between classes and the dispersion within the class Minimize data. To achieve this goal, we form two intra-class and inter-class dispersion matrices. Interdisciplinary matrix. The covariance matrix is the mean of each class and represents the distance and dispersion between classes. The class scattering matrix is the covariance of the data matrix of a class and indicates the degree of data scatter in that class. To achieve the maximum distance between the data of different classes and the greater proximity of the data of each class to each other, the matrix (scatter between classes / scatter within the class) should be maximized, which is done by calculating vectors and eigenvalues. They do this matrix. Then, like the Eigenface method, special vectors are used to transfer data to the space created by them. For each input image, we first move it to the desired subspace and then obtain the most similar image based on the similarity criterion. This method is better than the Eigen Face method because it differentiates between data of different classes and uses the class to which the data belongs to extract features and classification. has it. Also, the subspace to which the data is transferred in this method has smaller dimensions than the subspace produced by the Eigenface method, which makes the data more compact. To prevent singularity of covariance matrices within and between classes, a later reduction step is applied to the data before calculating them. The mathematical steps required to apply this method to the data are as follows.

a. Convert images to vectors and put them together to form a data matrix
b. Apply a later reduction step to the data.
c. Calculate the average of each class and the average of all classes.

d. Calculate the distance of each vector of a class from the average of the same class and form a scattering matrix within the class.
e. Calculate the average distance of each class from the average of all classes and form an interclass dispersion matrix.
f. Calculation of covariance of intra-class (SW) and interclass (SB) matrices.
g. Calculation of special vectors of the matrix 〖S_B S〗 _W ^ (- 1).
h. Data transfer using these special vectors.

The following should be done for the test image:

a. Convert image matrix to vector.
b. Calculate the difference of the data vector from the total average.
c. Move it below the formed space.
d. Compare with transferred vectors and determine the most similar vector.

The fisher face method uses linear separator analysis to transfer data to the desired subspace. The advantage of this method over the principal element's method is that the PCA method does not pay attention to the relationship between the data of a class (images belonging to a class) and classification is performed regardless of the class to which the data belongs. Be. On the other hand, if you do not pay attention to the data relationship within a class, it will lead to erroneous results. Figure 3.2 is the case in which the data are categorized by two methods, PCA and LDA. As can be clearly seen in the figure, the data were not properly classified in the PCA method. In the LDA method, the goal is to obtain a better space for expressing the data by considering the data appropriate to each class.

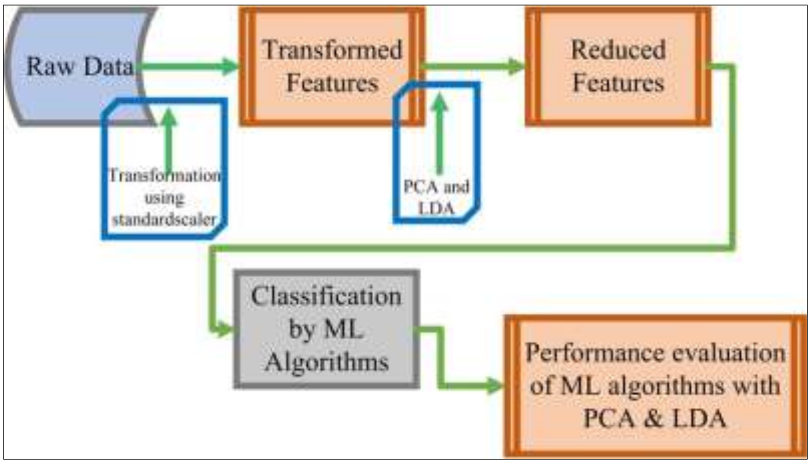

Figure 3.2: Point transfer using PCA, LDA [65].

As mentioned in the Fisher face method, the data are transferred to the space created by the base vectors obtained from the LDA method. The LDA method allows the data of different classes to be placed in the direction of maximum separation and the data within each class to be most closely related. The linear separation method, in addition to the mentioned advantages, also has several limitations, which include the following:

This method extracts a number of C-1 vectors, and if we find that we need more vectors, we must generate them using other methods. If the data distributions follow a non-Gaussian distribution, they cannot be separated since this classifier cannot produce nonlinear states.

This method fails in the case where the separator data is more in the variance and not in the center and cannot perform the separation properly.

3.5.1 Independent Component Analysis (ICA)

ICA analysis is a solution to the problem of blind resource separation in engineering. In the problem of blind separation of sources, it is assumed that the signal is obtained from the linear combination of several non-Gaussian and independent sources, and the purpose is to separate the source signals from the total signal. Similarly, in the case of face recognition, it can be assumed that each face is produced from the linear combination of several specific faces, so using the ICA extension, a face space can be formed for identification. ICA analysis, on the other hand, suggests a way to implement Barlow's redundancy theory for face recognition. Barlow sees systems redundancy as a structured knowledge of the environment, and believes that this redundancy makes it possible to encode objects that have high degrees of complexity. The use of redundancy in the human visual system has been studied by Ethics [66]. It is interesting to note that PCA analysis itself is a special case of ICA analysis and, assuming Gaussian sources (which is not an exact condition for images), consists of ICA expansion [67]. It should be noted that PCA analysis is only able to separate linear dependencies between components and there are dependencies of higher degrees in ICA expansion. The role of higher moment information in face recognition has been considered by many researchers [68].

In terms of image signal, second-order statistics do not include phase information. In his paper, Bartlett examines the effect of phase on facial images, arguing that if the phase of images changes, the nature of the image will change greatly [69]. Since higher moments are considered in ICA expansion and these moments can be considered as image phase information, Bartlett suggested the use of independent component analysis or ICA expansion to extract facial features. Based on his studies, he made the following claims in using the ICA extension:

a. If the dimensions of the feature vector are selected in the same ICA and PCA expansions, in most cases the vectors resulting from the ICA expansions will lead to better detection.
b. The sensitivity of ICA expansion vectors to dimensional change is less than that of PCA expansion vectors.
c. If we add noise to the feature vectors resulting from the two PCA and ICA expansions, the stability of the ICA expansion vectors in detection is greater than the PCA expansion.

Today, the veracity of Bartlett's claims is questioned, and the proximity - and in some scenarios - the decline of the ICA compared to machines using PCA or LDA extensions has led researchers to pay less and less attention to the ICA expansion over time [70][71].

3.5.2 Gabor Wavelet

In recent decades, many techniques for texture analysis have been proposed [72][73][74], including multi-channel filtering as an effective approach in various applications of image processing and machine vision. Based on studies by Zhang et al. [75], they found that the human visual system experiences the original image formed on the retina into a number of filtered images, each of which includes changes in light intensity over a narrow frequency range with Is a special direction. In other words, information processing is done by channels with a specific frequency band and direction. Mathematically, each of these channels is modeled with an intermediate filter called the Gabor filter. Researchers have concluded that the response of visual cell shocks is similar to that of a window-shaped Fourier transform with a Gaussian window [76]. The image, not the whole image, is calculated.

The Gabor wavelet calculates the properties of spatial localization, direction, spatial frequencies, and image phase dependence. It has been shown that the use of the Gabor wavelet is very appropriate when the aim is to find local and distinguishing features for image analysis and presentation. In the following, Gabor filter is introduced first and then Gabor features are described.

The Gabor wavelet equation is expressed in Formula 3.3 [77].

$$\varphi_{\mu,v}(z) = \frac{\|K_{\mu,v}\|^2}{\sigma^2} e^{\frac{\|K_{\mu,v}\|^2 \|z\|^2}{2\sigma^2}} \tag{3.3}$$

Where μ and v represent the direction and scale of Gabor's nuclei. z = (x, y), ‖.‖ The symbol of the soft operator and the wave vector k_ (μ, v) are calculated as follows:

$$k_{\mu,v} = k_v e^{i\emptyset_\mu} \tag{3.4}$$

3.6 PROPOPSED METHOD

Computer-based face recognition is a mature and reliable mechanism that is mainly used for many access control scenarios. In order to recognize and quantify facial characteristics in an image, computer face recognition features are used. Facial recognition technology may recognize human faces in images or videos, assess whether a face appears in two different images of the same person, or look for a face in a big database of existing images. Face recognition or authentication is mainly done using complete data from facial images. Therefore, face recognition uses imperfect faces and other facial manipulations such as rotation and zoom which are used as training and cognitive clues. The main topic of this research has been very important, therefore, the development of methods with acceptable accuracy and precision for face recognition remains the subject of many articles. There are various methods for face recognition that compare their accuracy with each other. Based on these results, computer-based face recognition models are a reliable and mature mechanism and are more suitable methods for resolving face-to-face diagnoses and predictions of face recognition and using this approach to solve similar problems is suggested. In this study, face recognition or authentication is mainly used with incomplete facial image data to improve the accuracy of face projections. The proposed method is also important, as it is imperative to reconstruct images with incomplete data and adapt them to other images. Therefore, this study considers the use of wavelet transforms based on fractal model and LSTM neural network. Mathematical methods are used to extract data from images using wavelets-based transforms. It has a substantial advantage over Fourier transforms in terms of temporal resolution, which means it can record both the frequency and position of the images. The fractal model starts with two main factors in the image segmentation phase including self-similarity and unique parts and also another operator called non-dimensionalityy. The fundamental concept behind fractals is that a basic process that undergoes infinitely many iterations develops into an extremely complicated process. Fractals look for the basic process behind the complicated process in an effort to simulate it. The majority of fractals function on the basis of a feedback loop. Wavelet transforms are also used for the initial reconstruction of faces. The wavelet transform (WT) may be used to evaluate data in time-frequency space, slow down on noise, and maintain the key elements of the original signals unaffected. WT has developed into a very efficient instrument for signal processing during the last 20 years. As start point, preparing image read in $\mathrm{N} \times \mathrm{N}$ measurement and go to $\mathrm{N}^2 \times 1$ measurement. One preparing set form in $\mathrm{N}^2 \times \mathrm{M}$ measurement that M is the quantity of test picture. Normal of picture set determined from condition (3.5).

$$\psi = \frac{1}{M} \sum_{i=1}^{M} \Gamma_i \tag{3.5}$$

In light of condition (1), ψ is normal picture, M is the quantity of pictures and Γ_i is picture vector. Relating Eigen faces with the most noteworthy Eigenvalues hold which these Eigen faces characterize face space. Eigen-space framed by picture projection to picture space which caused Eigen faces. Eigenfaces is a method for face identification and detection that uses the variance of faces in a collection of face images to encode and decode a face in a machine learning approach without the entire information, minimizing computational and space complexity. Thus, vector weight are determined. Picture measurement tune for properties perception and picture will be upgrade in pre-preparing step. At that point picture weight vector and faces weight vector think about in picture information base. At that point normal face determined and deduct from each current picture in preparing set. A Matrix will be worked by the consequences of deduction activity. Correlation between each picture and picture normal determined as condition (3.6).

$$\phi_i = \Gamma_i - \Psi, \qquad i = 1,2, \dots, M \tag{3.6}$$

In condition (2), ϕ_i is the contrast among picture and picture normal. Acquired network from deduction activity or A lattice, duplicate in its rendered lastly C covariance framework structure with condition (3.7).

$$\mathrm{C} = \mathrm{A}^{\mathrm{T}}\mathrm{A} \tag{3.7}$$

In light of condition (3), A lattice framed from contrast between vectors. As an illustration it very well may be brought up to $A = [\phi_1, \phi_2, \phi_3, \dots \phi_M]$. C Matrix measurement resembles $N \times M$. Then again, when A level is equivalent to M, just M from N number of Eigenvalue is equivalent to non-zero worth. At that point Eigenvalues of covariance framework determined and Eigen faces work by some prepared pictures deduct to number of Eigenvalues class. The quantity of Eigen-vector classes are the absolute accessible appearances in picture. Chosen sets of Eigen-vectors by A grid duplicate because of measurement decrease of Eigen face. Eigen-vectors of more modest Eigenvalues match to more modest changes in covariance grid. The variance components of the covariance matrix and the eigenvalues continue to indicate the variance magnitude in the direction of the x-axis and y-axis, respectively, and the variance magnitude in the direction of the direction of the data's greatest distribution, respectively. Different highlights of picture keep up. The quantity of Eigen-vectors relies upon picture precision in information base. A portion of the chose Eigenvalues are Eigen faces. At the point when Eigen faces got, pictures at information base put away to Eigen face space and picture loads in that projection of picture space. Eigen-coefficients contrasted for characterize a picture and Eigen-coefficient from picture information base. At that point Eigen face work in picture. Euclidian distance determined between Eigen face of picture and put away Eigen faces in past advance. The Euclidean distance may be used to determine the separation between any two points in two dimensions as well as to compute the absolute separation between points in N-dimensional space. Smaller values represent more comparable faces when identifying faces. Expected individual perceive as an individual that Euclidian distance is lower than edge esteem in Eigen face data set. Assuming the entirety of the Euclidian distance are greater than limit esteem, there are no acknowledgment of face in picture and it will be neglected. We attempt to separate highlights like roundabout face area, eyes, and edge of cheeks. In the entirety of the component extraction if there are less highlights with the best outcome, it will show that the element extraction was enhance and can be utilized in genuine.

It is necessary to propose an algorithm for classifying and recognizing imperfect faces. In general, it can be argued that LSTM is exactly the opposite of deep learning methods and alternative classification methods such as Support Vector Machine and Naïve Bayesian. Naive Bayes algorithms are frequently used in sentiment analysis, spam filtering, recommendation systems, etc. The necessity that predictors be independent is their major drawback even though that they are quick and simple to implement. The LSTM algorithm can employ nonlinear activation functions such as sigmoid or sinusoidal or non-derivative activation functions as well as using a linear function to activate cells or neurons in the hidden layer because of its high flexibility. A deep neural network is needed to learn complicated data sets with high accuracy, and this network is generated by stacking numerous layers of neurons with a non-linear activation function. A sigmoid activation function for a neuron ensures that the output of this component will always range between 0 and 1. The sigmoid is a non-linear function, hence this unit's output would also be a non-linear function of the weighted sum of its inputs. By default, LSTM has an equation in general mode such as equation (3.8).

$$y(p) = \sum_{j=1}^{m} \beta_i \beta_j g(\sum_{i=1}^{n} w_{i,j} x_i + b_j) \tag{3.8}$$

According to this equation, β_i represents the weights between the input layer and the hidden layer, and β_j represents the weights between the output layer and the input layer. b_j is the threshold value of neurons in the hidden layer, or bias. $g(\ ...)$ is the transition or actuator function. $w_{i,j}$ is the input layer weights and b_j is the bias which are randomly assigned. Activation function $g(\ ...)$ is assigned at the start of the number of input layer neurons or n and the number of hidden layer neurons or m. According to this information, if the known parameters in the overall balancing are combined and recalibrated, the output layer will be like equation (3.9).

$$H(w_{i,j}, b_j, x_i) = \begin{bmatrix} g(w_{1,1}x_1 + b_1) & \cdots & g(w_{1,m}x_m + b_m) \\ g(w_{n,1}x_n + b_1) & \cdots & g(w_{n,m}x_m + b_m) \end{bmatrix} \quad \text{and} \quad y = H\beta \tag{3.9}$$

The main goal in all models of training-oriented algorithms is to minimize the error as much as possible. The fundamental phase in machine learning is model training, which results in a functioning model that can subsequently be verified, tested, and deployed. How well a model performs during training ultimately determines how well it will function when it is implemented in an application for end users. The output error function y_p is obtained by the actual output y_{main} in LSTM, which can be represented by two training sections namely $\sum_k^s (y_{main} - y_p)$ and testing sections namely $\left\| \sum_k^s (y_{main} - y_p)^2 \right\|$. For both functions, the output y_p obtained by the actual output y_{main} must be equal to y_p. An unknown parameter is specified when this equation is executed and the results are satisfied. The matrix H can be a matrix that is very unlikely, meaning that the amount of data in the training set phase may not be equal to the total number of data attributes. So reversing $[H]$ and finding weights or β will be a major challenge. To overcome this challenge in LSTM, a matrix called Moore Penrose is used that can develop approximate inverse matrix computation that can perform dimensionality selection and feature extraction operations with classification with high accuracy and incredible speed compared to other methods. The Moore-Penrose pseudo inverse is an extension of the matrix inverse when the matrix is not invertible. The Moore-Penrose pseudo inverse is identical to the matrix inverse if A is invertible. Even if A is not invertible, the Moore-Penrose pseudo inverse is defined. When applied to square singular matrices and rectangular matrices, a generalized inverse is a generalization of the definition of an inverse. There are several definitions of generalized inverses, all of which, when the matrix is square and non-singular, reduce to the standard inverse. Furthermore, it helps to simplify the investigation of solutions to arbitrary systems of linear equations and linear least squares problems. Using the Moore Penrose matrix, β^* is the output matrix and H^* is the generalized inverse Penrose matrix of H. Therefore, due to the optimization of the LSTM, the problem of output weights in the LSTM was solved as $B^* = H^*$ which became the MPM-LSTM. In general, MPM-LSTM becomes a chain of repeating modules over time in the training phase. MPM-LSTM will be able to work like a conveyor that is to add or subtract information to neurons. No weight updating operations are performed during training with MPM-LSTM method unlike deep learning structures and other classification models such as Support Vector Machine or Naïve Bayesian.

MPM-LSTM can specify attributes. Since the segmented parts which applied by fractal model have been entered the feature extraction phase and the main features include brightness and edges intensity, which have also been reduced to noise and image enhancements in the earlier stages, the dimension reduction and feature selection operation are performed with MPM-LSTM and represent the exact face recognition from imperfect data.

4. EXPERIMENTAL RESULTS AND DISCUSSION

In the following, the face recognition results are described. First of all the dataset will describe and the result of the proposed method will present.

4.1 DATA SET

Generally, the proposed approach will be tested with each section's description on a series of images from the standard dataset VGG Face, and examines its results. Initially, an image is required to be programmed as input. The input image is as Figure 4.1.

Figure 4.1: Input Image.

Following is the image segmentation operation. During image segmentation, the image components are removed. Segmentation is a crucial step in the image recognition process because it isolates the items of interest for future processing, such as description or identification. The categorization of image pixels sometimes involves segmenting the image. This operation can display prominent features. Fractal model is used in segmentation using two components of brightness and edge intensity with two operator means self-similarity and unique parts. In fact, the image features can be distinguished and extracted by segmenting the image. So, morphology is an important topic in this part. Hence, the four main features of the face, including the face, eye, lip and cheek area, are used to identify the face in the image. The output of this section for segmentation is as Figure 4.2.

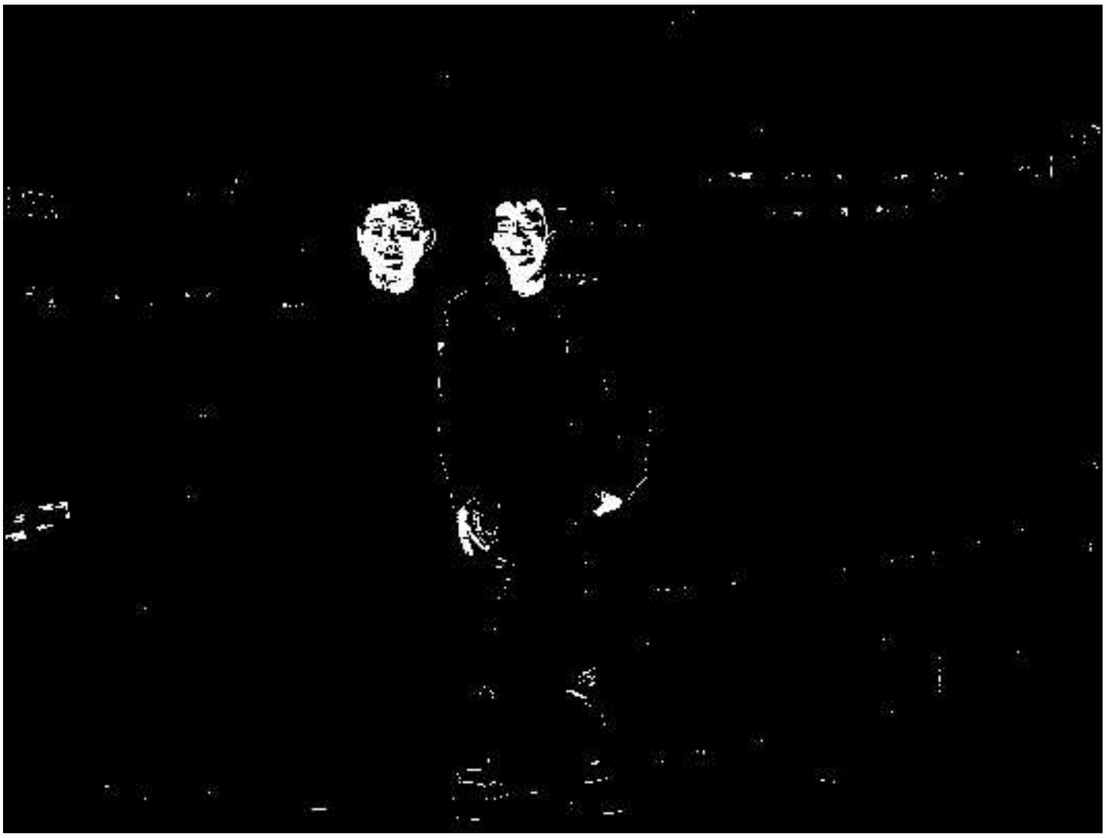

Figure 4.2: Fractal Model-Based Image Segmentation with Feature Extraction.

Once the segmentation is complete and the features are specified, the feature dimension reduction, feature selection and extraction operations are required along with image reconstruction. Face detection frequently uses the OpenCV approach. The AdaBoost technique is used as the face detector after first extracting the feature images from a large sample set using the face Hear features. The operation is based on the wavelet transform method. After identifying the face area in the image, a matrix is formed in that area, identifying the edge of that area and identifying features such as the face area, eyes, lips and cheeks that work based on morphology. The output of this section is as figure 4.3.

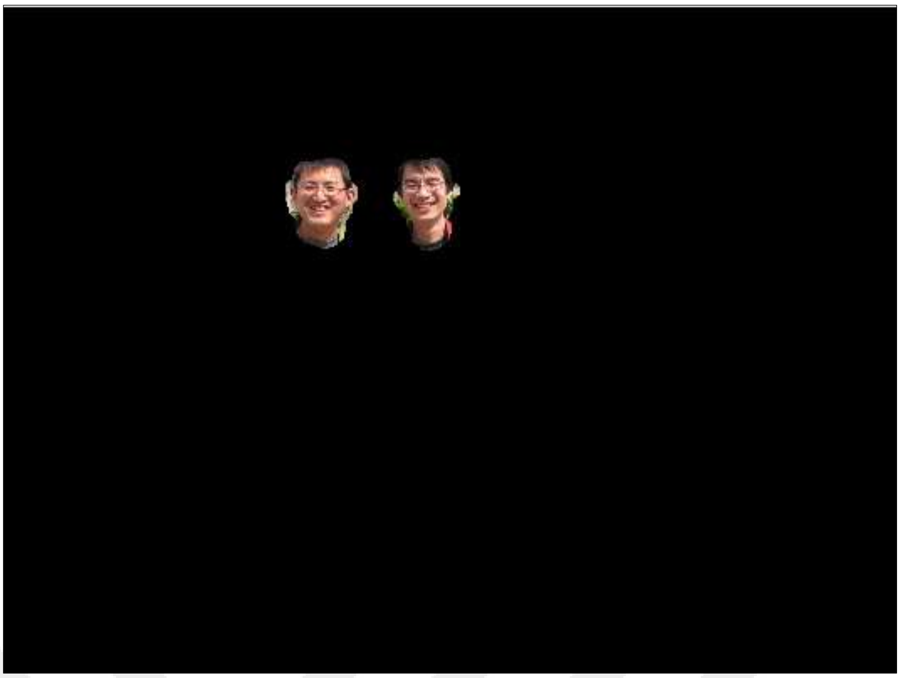

Figure 4.3: Image Reconstruction with Wavelet Transform.

It should be noted that the wavelet transform method follows a randomized training method, but is not capable of accurately identifying the face area after reconstruction. This has occurred on one image, but if more data is entered, more accurate identification is needed. The technique of turning unprocessed raw data into numerical features that may be used for processing while maintaining the data from the original data set is known as feature extraction. It produces superior outcomes compared to machine learning that is applied directly to the raw data. So, in another image, there may not be a face at all, or it may not be clear. Therefore, the use of a learning system after segmentation, feature extraction, and reconstruction is needed. This research considers the use of an improved method called MPM-LSTM. In fact, the MPM-LSTM has been used for accurate face recognition training and for imperfect face data recognition. The results of this section are shown in figure 4.4.

Figure 4.4: Training and Testing with MPM-LSTM.

It is observed that imperfect face identification in reliable mode is performed in Fig (4) in a few simple steps based on the proposed approach. In the following, another image is trained and tested in the form of Fig (4.5) to (4.8).

Figure 4.5: Results of Proposed Method in four Steps.

Figure 4.6: Outcomes of the four-Step Recommended Technique.

Figure 4.7: Outcomes of the Proposed Four-Step Technique.

Figure 4.8: Findings from the Four-Step Recommended Method.

Overall, the training and test results are based on 100 images from the VGG Face dataset for which the ROC chart will be drawn. The ROC diagram, known as the "receiver operating characteristic curve", is a quantitative evaluation method that is curved. The Receiver Operating Characteristics (ROC) curve is a binary classifier assessment tool that enables us to determine how well a facial recognition model performs as the discrimination threshold increases. It is mainly used in image processing, data mining and signal processing projects. This curve shows the differences between data identification methods. In this curve, the real purpose is to accurately evaluate the pathological or error ability of a system, thus using criteria based on decisions. The output of the ROC curve is as shown in Fig (4.9), and the area under the curve, or AUC is 0.86, which is less than 1 indicating the standard and appropriate fit of the proposed approach.

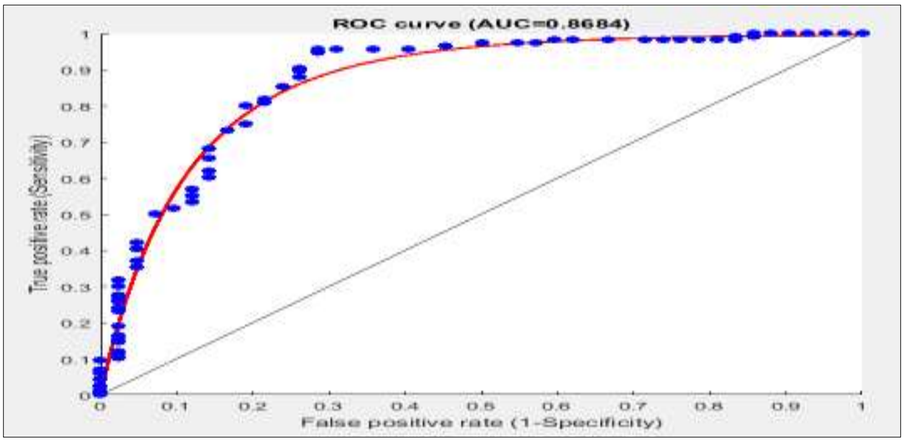

Figure 4.9: Proposed Method ROC Curve.

The results of the evaluation criteria include Mean Squares Error (MSE), accuracy, sensitivity, and specificity after using 100 images of the VGG Face dataset to identify faces with incomplete data with the proposed approach in Table (4.1) and comparison of proposed method and recent method represent in the Table (4.2) and Fig (4.10). VGG Face is a dataset that contains 2.6 million images of faces from 2,622 individuals and is used to generate face recognition software. A text file with URLs for images and accompanying facial detections is linked to each identity. Before downloading the data, please carefully read the agreement file. The VGG Face Descriptor website contains models that have been pre-trained using this data.

Table 4.1: Evaluation Criteria of Proposed Method

Specificity (%)	Sensitivity (%)	Accuracy (%)	MSE
84.16 %	91.33 %	99.58 %	0.09

Table 4.2: Comparison of Proposed and Recent Methods.

Method	Accuracy
L-SVM	98 %
L-SVM-Wo	98.5 %
k-SVM	95 %
k-SVM-Wo	94.5 %
CS	98.1 %
CS-Wo	98.3 %
Proposed Method (FW-MPM-LSTM)	99.58 %

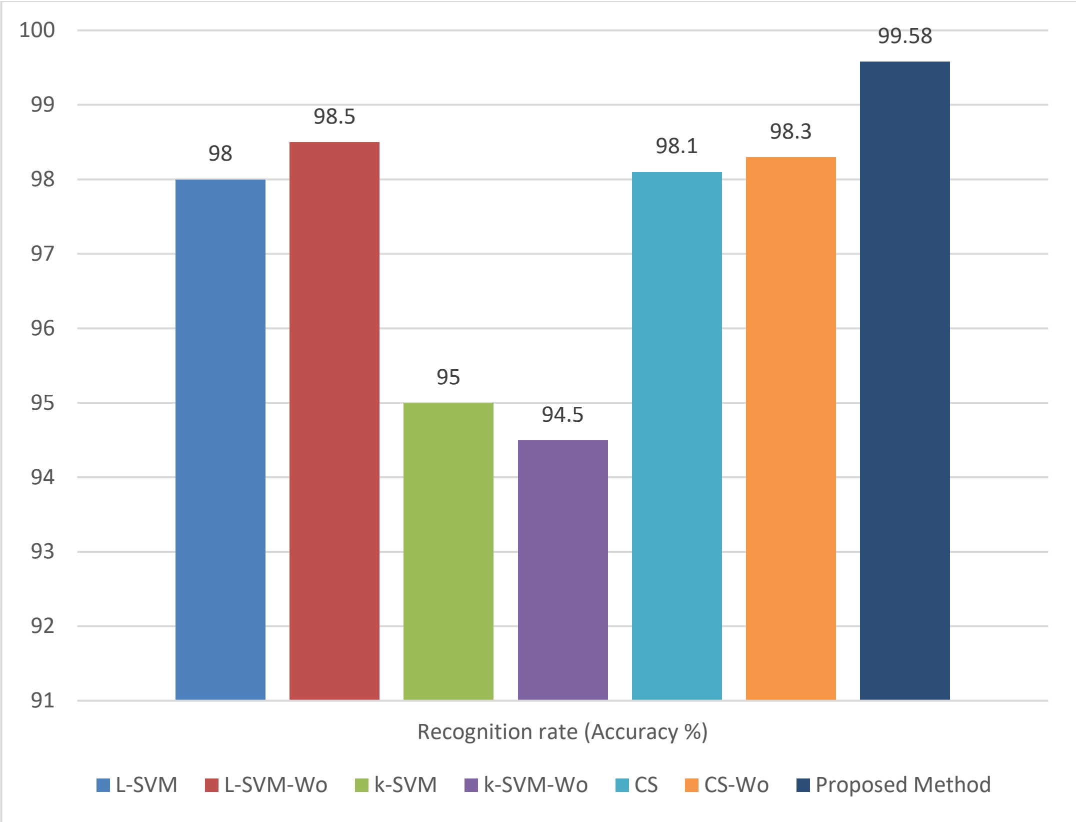

Figure 4.10: Comparison of Proposed and Recent Methods.

Form the results it can be understand that the combination of the MPM method with the deep learning that used the LSTM give the high accuracy and high performance that other methods that mentioned in the table. Also, the segmentation method increases the sensitivity, specificity and accuracy of the recognition office in crowded scene.

5. CONCLUSIONS AND FUTURE WORK

5.1 CONCLUSIONS

Conventional face recognition methods are divided into three main groups. Appearance, modeling and hybrid methods. Each of these methods can evaluate and process all or part of a person's face. The use of appearance methods to solve face recognition problems has been the most popular among researchers compared to model methods. Most detection machines use the brightness of the pixels as input data. However, the brightness of the face may change somewhat due to the rotation of the face and the change in the brightness of the environment. Recently, facial recognition research has focused on facial expressions that are resistant to changes in brightness and facial expressions, leading to the development of a powerful texture descriptor called the Local Binary Pattern (LBP). The LBP method is one of the best local appearance techniques that expresses the texture, which has been widely used in various applications in recent years.

Face recognition from imperfect data and their matching is an important issue in the principles of image processing and machine vision. In some shooting situations, some of the important components of a person's face may not be included in the images, and this may complicate the processing which used to identify faces and future goals such as face detection and matching, identifying people's crime, detecting anomalies, and etc. Therefore, providing an optimal and efficient method with maximum accuracy can be an interesting problem in the field of face recognition from imperfect data. The proposed approach is to identify and reconstruct four features that include face, lips, eyes and cheeks. Based on these four features, pre-processing is performed first, followed by segmentation, feature extraction and image reconstruction using the fractal model and wavelet transform.

The LSTM neural network is then used to train and test imperfect face data, but due to the weaknesses of this neural network, the Moore Penrose matrix is used to improve it and the method will be MPM-LSTM. The simulation results in face recognition for all the features mentioned in the imperfect data show that the proposed approach called FW-MPM-LSTM has a higher performance accuracy than its recent methods. The recognition rate for proposed method was obtained 99.58, while this percentage for L-SVM, L-SVM-Wo, K-SVM, K-SVM-Wo, CS, CS-Wo were obtained 98, 98.5, 95, 94.5, 98.1, and 98.3 respectively.

5.2 FUTURE WORK

In the future, we can focus our efforts on developing various metaheuristic optimization methods, such as ant colony optimization and gray wolf optimization, with the end goal of lowering the feature dimension while simultaneously raising the accuracy.

REFERENCES

[1] S. M. Andersen, C. A. Carlson, M. A. Carlson, and S. D. Gronlund, “Individual differences predict eyewitness identification performance,” Pers. Individ. Dif., vol. 60, pp. 36–40, 2014.

[2] J. W. Tanaka, M. D. Kaiser, S. Hagen, and L. J. Pierce, “Losing face: impaired discrimination of featural and configural information in the mouth region of an inverted face,” Attention, Perception, Psychophysics., vol. 76, no. 4, pp. 1000–1014, 2014.

[3] Z. Li, J. Liu, J. Tang, and H. Lu, “Robust structured subspace learning for data representation,” IEEE Trans. Pattern Anal. Mach. Intel., vol. 37, no. 10, pp. 2085–2098, 2015.

[4] C. Ding and D. Tao, “Pose-invariant face recognition with homography-based normalization,” Pattern Recognit., vol. 66, pp. 144–152, 2017.

[5] C. Ding and D. Tao, “Robust face recognition via multimodal deep face representation,” IEEE Trans. Multimed., vol. 17, no. 11, pp. 2049–2058, 2015.

[6] T. Zhang, J. Li, W. Jia, J. Sun, and H. Yang, “Fast and robust occluded face detection in ATM surveillance,” Pattern Recognit. Lett., vol. 107, pp. 33–40, 2018.

[7] D. Marčetić and S. Ribarić, “Deformable part-based robust face detection under occlusion by using face decomposition into face components,” in 2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1365–1370, 2016.

[8] A. Jadhav, A. Jadhav, T. Ladhe, and K. Yeolekar, “Automated attendance system using face recognition,” Int. Res. J. Eng. Technol., vol. 4, no. 01, pp. 1467–1471, 2017.

[9] S. Chintalapati and M. V Raghunadh, “Automated attendance management system based on face recognition algorithms,” in 2013 IEEE International conference on computational intelligence and computing research, pp. 1–5, 2013.

[10] N. M. Kutty and S. Mathai, “Face recognition-a tool for automated attendance system,” Int. Journals Adv. Res. Comput. Sci. Softw. Eng., vol. 7, no. 6, pp. 334–336, 2017.

[11] P. S. K. Patra, G. D. Kumar, R. BalaMurali, and G. Murugan, “Multimodal biometric system over Unimodal biometric system”.

[12] S. Atawneh, A. Almomani, H. Al Bazar, P. Sumari, and B. Gupta, “Secure and imperceptible digital image steganographic algorithm based on diamond encoding in DWT domain,” Multimed. Tools Appl., vol. 76, no. 18, pp. 18451–18472, 2017.

[13] A. M. Bukar and H. Ugail, “Automatic age estimation from facial profile view,” IET Comput. Vis., vol. 11, no. 8, pp. 650–655, 2017.

[14] H. Ugail and A. Al-Dahoud, “Is gender encoded in the smile? A computational framework for the analysis of the smile driven dynamic face for gender recognition,” Vis. Comput., vol. 34, no. 9, pp. 1243–1254, 2018.

[15] B. Olk and A. M. Garay-Vado, “Attention to faces: Effects of face inversion,” Vision Res., vol. 51, no. 14, pp. 1659–1666, 2011.

[16] J. Taubert, D. Apthorp, D. Aagten-Murphy, and D. Alais, “The role of holistic processing in face perception: Evidence from the face inversion effect,” Vision Res., vol. 51, no. 11, pp. 1273–1278, 2011.

[17] J. J. Richler, M. L. Mack, T. J. Palmeri, and I. Gauthier, “Inverted faces are (eventually) processed holistically,” Vision Res., vol. 51, no. 3, pp. 333–342, 2011.

[18] S. Saypadith and S. Aramvith, “Real-time multiple face recognition using deep learning on embedded GPU system,” in 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), pp. 1318–1324, 2018.

[19] F. Deeba, H. Memon, F. A. Dharejo, A. Ahmed, and A. Ghaffar, “LBPH-based enhanced real-time face recognition,” Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 5, 2019.

[20] M. D. Nisha, “Improving the recognition of faces using LBP and SVM optimized by PSO technique,” Int. J. Eng. Dev. Res., vol. 5, no. 4, pp. 297–303, 2017.

[21] B. O’Connor and K. Roy, “Facial recognition using modified local binary pattern and random forest,” Int. J. Artif. Intell. Appl., vol. 4, no. 6, pp. 25, 2013.

[22] S. M. Bah and F. Ming, “An improved face recognition algorithm and its application in attendance management system,” Array, vol. 5, pp. 100014, 2020.

[23] P. M. Kumar, U. Gandhi, R. Varatharajan, G. Manogaran, and T. Vadivel, “Intelligent face recognition and navigation system using neural learning for smart security in Internet of Things,” Cluster Comput., vol. 22, pp. 7733–7744, 2019.

[24] S. Ahmed, M. Frikha, T. D. H. Hussein, and J. Rahebi, “Face Recognition System using Histograms of Oriented Gradients and Convolutional Neural Network based on with Particle Swarm Optimization,” in 2021 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), pp. 1–5, 2021.

[25] S. Ahmed, M. Frikha, T. D. H. Hussein, and J. Rahebi, “Optimum Feature Selection with Particle Swarm Optimization to Face Recognition System Using Gabor Wavelet Transform and Deep Learning,” Biomed Res. Int., vol. 2021, 2021.

[26] S. Allagwail, O. S. Gedik, and J. Rahebi, “Face recognition with symmetrical face training samples based on local binary patterns and the Gabor filter,” Symmetry (Basel)., vol. 11, no. 2, pp. 157, 2019.

[27] B.-B. Benuwa, B. Ghansah, and E. K. Ansah, “Kernel based locality–sensitive discriminative sparse representation for face recognition,” Sci. African, vol. 7, pp. e00249, 2020.

[28] J. Chen, R. D. J. Samuel, and P. Poovendran, “LSTM with bio inspired algorithm for action recognition in sports videos,” Image Vis. Comput., vol. 112, pp. 104214, 2021.

[29] J. Fernández-Montoya, C. Avendaño, and P. Negredo, “The glutamatergic system in primary somatosensory neurons and its involvement in sensory input-dependent plasticity,” Int. J. Mol. Sci., vol. 19, no. 1, pp. 69, 2018.

[30] S. K. Esser, R. Appuswamy, P. Merolla, J. V Arthur, and D. S. Modha, “Backpropagation for energy-efficient neuromorphic computing,” Adv. Neural Inf. Process. Syst., vol. 28, pp. 1117–1125, 2015.

[31] M. Turčaník and M. Javurek, “Hash function generation by neural network,” in 2016 New Trends in Signal Processing (NTSP), pp. 1–5, 2016.

[32] M. Turčaník, “Hash function generation based on neural networks and chaotic maps,” in 2017 Communication and Information Technologies (*KIT)*, pp. 1–5, 2017.

[33] N. Abdoun, S. El Assad, R. Assaf, O. Déforges, M. Khalil, and S. Belghith, “Design and implementation of robust Keyed Hash functions based on Chaotic Neural Network,” 2018.

[34] B. Karlik and A. V. Olgac, “Performance analysis of various activation functions in generalized MLP architectures of neural networks,” Int. J. Artif. Intell. Expert Syst., vol. 1, no. 4, pp. 111–122, 2011.

[35] B. Xu, N. Wang, T. Chen, and M. Li, “Empirical evaluation of rectified activations in convolutional network,” arXiv Prepr. arXiv1505.00853, 2015.

[36] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” J. Mach. Learn. Res., vol. 15, no. 1, pp. 1929–1958, 2014.

[37] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv Prepr. arXiv1412.6980, 2014.

[38] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in Proceedings of the thirteenth international conference on artificial intelligence and statistics, pp. 249–256, 2010.

[39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” Commun. ACM, vol. 60, no. 6, pp. 84–90, 2017.

[40] H.-C. Shin *et al.*, “Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning,” *I*EEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1285–1298, 2016.

[41] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826, 2016.

[42] K. R. Konda, “Unsupervised Relational Feature Learning for Vision.” Univ.-Bibliothek Frankfurt am Main, 2016.

[43] I. K. Sethi and A. K. Jain, Artificial neural networks and statistical pattern recognition: old and new connections. Elsevier, 2014.

[44] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Neural networks for perception*, Elsevier, pp. 65–93, 1992.

[45] R. Miikkulainen *et al.*, "Evolving deep neural networks," in Artificial intelligence in the age of neural networks and brain computing, Elsevier, pp. 293–312, 2019.

[46] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural networks, vol. 61, pp. 85–117, 2015.

[47] L. Torrey and J. Shavlik, "Transfer learning," in Handbook of research on machine learning applications and trends: algorithms, methods, and techniques, IGI global, pp. 242–264, 2010.

[48] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in International conference on artificial neural networks, pp. 270–279, 2018.

[49] H. Ravishankar et al., "Understanding the mechanisms of deep transfer learning for medical images," in Deep learning and data labeling for medical applications, Springer, pp.188–196, 2016.

[50] X. Yin, X. Yu, K. Sohn, X. Liu, and M. Chandraker, "Feature transfer learning for face recognition with under-represented data," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5704–5713, 2019.

[51] K. Gopalakrishnan, S. K. Khaitan, A. Choudhary, and A. Agrawal, "Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection," Constr. Build. Mater., vol. 157, pp. 322–330, 2017.

[52] Y. Yu, H. Lin, J. Meng, X. Wei, H. Guo, and Z. Zhao, "Deep transfer learning for modality classification of medical images," Information, vol. 8, no. 3, pp. 91, 2017.

[53] A. Boschetti, N. Adami, R. Leonardi, and M. Okuda, "Image coding with face descriptors embedding," in 2011 18th IEEE International Conference on Image Processing, pp. 3665–3668, 2011.

[54] A. K. Jain and S. Z. Li, Handbook of face recognition, vol. 1. Springer, 2011.

[55] C. Chen, Z. Hua, R. Zhang, G. Liu, and W. Wen, "Automated arrhythmia classification based on a combination network of CNN and LSTM," Biomed. Signal Process. Control, vol. 57, pp. 101819, 2020.

[56] J. R. Beattie and F. W. L. Esmonde-White, "Exploration of principal component analysis: deriving principal component analysis visually using spectra," Appl. Spectrosc., vol. 75, no. 4, pp. 361–375, 2021.

[57] B. M. S. Hasan and A. M. Abdulazeez, "A review of principal component analysis algorithm for dimensionality reduction," J. Soft Comput. Data Min., vol. 2, no. 1, pp. 20–30, 2021.

[58] F. Kherif and A. Latypova, "Principal component analysis," in Machine Learning, Elsevier, 2020, pp. 209–225.

[59] S. X. Wu, H.-T. Wai, L. Li, and A. Scaglione, "A review of distributed algorithms for principal component analysis," *Proc. IEEE*, vol. 106, no. 8, pp. 1321–1340, 2018.

[60] D. Gray, K. Yu, W. Xu, and Y. Gong, "Predicting facial beauty without landmarks," in *European Conference on Computer Vision*, pp. 434–447, 2010.

[61] V. AHLAWAT, "FACE RECOGNITION USING INFRARED IMAGING." 2019.

[62] I. Adjabi, A. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed, "Past, present, and future of face recognition: A review," Electronics, vol. 9, no. 8, pp. 1188, 2020.

[63] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," IEEE Trans. Pattern Anal. Mach. Intell., no. 7, pp. 711–720, 1997.

[64] S. Roopashree, J. Anitha, T. R. Mahesh, V. V. Kumar, W. Viriyasitavat, and A. Kaur, "An IoT based Authentication System for Therapeutic Herbs measured by Local Descriptors using Machine Learning Approach," Measurement, pp. 111484, 2022.

[65] G. T. Reddy *et al.*, "Analysis of dimensionality reduction techniques on big data," IEEE Access, vol. 8, pp. 54776–54788, 2020.

[66] H. Park, R. A. A. Ince, P. G. Schyns, G. Thut, and J. Gross, "Representational interactions during audiovisual speech entrainment: Redundancy in left posterior superior temporal gyrus and synergy in left motor cortex," PLoS Biol., vol. 16, no. 8, pp. e2006558, 2018.

[67] A. de Borman, P.-A. Absil, R. El Tahry, and A. Kumar, "Estimation of seizure onset zone from ictal EEG using independent component analysis and source imaging." 2021.

[68] H. K. Palo, S. Sahoo, and A. K. Subudhi, "Dimensionality reduction techniques: Principles, benefits, and limitations," Data Anal. Bioinforma. A Mach. Learn. Perspect., pp. 77–107, 2021.

[69] M. Leszczynski, L. Chaieb, T. Staudigl, S. J. Enkirch, J. Fell, and C. E. Schroeder, "Neural activity in the human anterior thalamus during natural vision," Sci. Rep., vol. 11, no.1, pp. 1–13, 2021.

[70] D. K. Choubey, M. Kumar, V. Shukla, S. Tripathi, and V. K. Dhandhania, "Comparative analysis of classification methods with PCA and LDA for diabetes," Curr. Diabetes Rev., vol. 16, no. 8, pp. 833–850, 2020.

[71] S. Karthick, S. Selvakumarasamy, C. Arun, and P. Agrawal, "Automatic attendance monitoring system using facial recognition through feature-based methods (PCA, LDA)," Mater. Today, Proc., 2021.

[72] R. Hammouche, A. Attia, S. Akhrouf, and Z. Akhtar, "Gabor filter bank with deep autoencoder based face recognition system," Expert Syst. Appl., pp.116743, 2022.

[73] C. Li, Y. Huang, W. Huang, and F. Qin, "Learning features from covariance matrix of gabor wavelet for face recognition under adverse conditions," Pattern Recognit., vol.119, pp. 108085, 2021.

[74] R. Mohammadian Fini, M. Mahlouji, and A. Shahidinejad, "Performance improvement in face recognition system using optimized Gabor filters," Multimed. Tools Appl., pp.1–34, 2022.

[75] Y. Zhang, W. Zhou, and J. Xu, "A dynamic range adjustable inverse tone mapping operator based on human visual system," Vis. Comput., pp.1–15, 2022.

[76] C. R. Meixner *et al.*, "Hybrid-shimming and gradient adaptions for improved pseudo-continuous arterial spin labeling at 7 Tesla," Magn. Reson. Med., vol.87, no.1, pp. 207–219, 2022.

[77] S. K. Rajeev, M. P. Rajasekaran, G. Vishnuvarthanan, and T. Arunprasath, "A biologically-inspired hybrid deep learning approach for brain tumor classification from magnetic resonance imaging using improved gabor wavelet transform and Elmann-BiLSTM network," Biomed. Signal Process. Control, vol. 78, pp. 103949, 2022.