REPUBLIC OF TURKEY
FIRAT UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND
APPLIED SCIENCE

LOCAL RECEPTIVE FIELDS BASED EXTREME
LEARNING MACHINE FOR FACE RECOGNITION

ARAS MASOOD ISMAEL

Master Thesis
Department: Software Engineering
Supervisor: Prof. Dr. Abdulkadir ŞENGÜR

FEBRUARY-2018

REPUBLIC OF TURKEY

FIRAT UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE

LOCAL RECEPTIVE FIELDS BASED EXTREME LEARNING
MACHINE FOR FACE RECOGNITION

MASTER THESIS

Aras Masood ISMAEL

Supervisor: Prof. Dr. Abdulkadir ŞENGÜR

Other Jury Members: Assoc. Prof. Dr. Murat KARABATAK

: Assist. Prof. Dr. Ömer F. ALÇİN

FEBRUARY-2018

ACKNOWLEDGEMENT

CONTENTS

ABSTRACT

Local Receptive Fields Based Extreme Learning Machine for Face Recognition

In the recent years, many face recognition methods were developed and applied in the image processing applications, such as artificial neural networks (ANN), convolutional neuron network (CNN), Gaussian mixtures and so on. However, each of them is suffering from some issues like the local minima, intensive human intervention rate, and slow convergence. CNN generally uses the back propagation (BP) learning procedure, which has very long training periods. In addition, back-propagation generally tends to reach a local minimum. While each of the above methods needs many iterations to reduce the error rate in the training section, we apply extreme learning machine, which it runs throw pooling map with Local receptive fields, in this case, the proposed method can get the result just in one iteration. In addition, ELM proposed to alleviate these drawbacks of the back-propagation method. The Extensive experiments were conducted on three face datasets namely Caltech, UFI, and CBCL. The obtained results are encouraging and compared with several other results previously reported. Testing accuracy in Caltech face dataset is 98.15%, also in the CBCL dataset is 98.34% and in UFI face dataset is 66.11%. Based on the above result the extreme learning machine based on local receptive fields can have more advantage among previous methods used for face recognition.

Keywords: Deep extreme learning machine, face recognition, image processing.

ÖZET

Yüz Algılama İçin Yerel Algılayıcı Alanlara Dayalı Aşırı Öğrenme Makinesi

Son yıllarda birçok yüz tanıma yöntemi geliştirilmiş ve yapay sinir ağları (ANN), konvolusyonel nöron ağı (CNN), Gauss karışımları ve benzeri gibi görüntü işleme uygulamalarında uygulanmıştır. Bununla birlikte, her biri yerel minimum, yoğun insan müdahale oranı ve yavaş yakınsama gibi bazı konulardan muzdarip. CNN genellikle çok uzun eğitim sürelerine sahip olan geri yayılım (BP) öğrenme prosedürünü kullanır. Buna ek olarak, geri yayılım genellikle bir yerel minimuma ulaşma eğilimindedir. Yukarıdaki yöntemlerin her biri, eğitim bölümündeki hata oranını düşürmek için birçok yineleme gerektirse de, aşırı öğrenme makinesini uyguladığımızda, bu da yerel algılayıcı alanları ile eşleme haritasını atamaktadır; bu durumda, önerilen yöntem sadece tek bir iterasyonda sonuçları elde edebiliyor. Buna ek olarak ELM, geri yayılım yönteminin bu dezavantajlarını hafifletmeyi önermiştir. Ayrıca, Caltech, UFI ve CBCL olmak üzere üç yüz veri setinde kapsamlı deneyler yapılmıştır. Elde edilen sonuçlar cesaret vericidir ve daha önce bildirilen birkaç başka sonuç ile de karşılaştırılmıştır. Caltech yüz veri kümesinde test doğruluğu %98.15, CBCL veri kümesinde %98.34 ve UFI yüz veri kümesinde %66.11'dir. Bu nedenle, yerel algılayıcılı alanlarına dayanan aşırı öğrenme makinesinin yüz tanımada kullanılan önceki yöntemler arasında daha fazla avantaj sağlayabileceğini görebiliriz.

Anahtar Kelimeler: Derin aşırı öğrenme makinesi, yüz tanıma, görüntü işleme

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ANN	**:**	Artificial Neural Network
ATM	**:**	Automated Teller Machine
BP	**:**	Back-propagation
CNN	**:**	Conventional Neural Network
ELM	**:**	Extreme Learning Machine
FA	**:**	Face Recognition
FS-LBP	**:**	Face Specific- Local Binary Pattern
LDN	**:**	Local directional number
LDP	**:**	Local Derivative Patterns
LDP-HS	**:**	Local Derivative Pattern- Histogram Sequences
LGBPHS	**:**	Local Gabor Binary Pattern Histogram Sequences
LRF	**:**	Local Receptive fields
LTP	**:**	Local Ternary Patterns
MPLCALT	**:**	Multi-linear Principal Component Analysis Local Ternary
PCA	**:**	Principal Component Analysis
PIN	**:**	Personal identification numbers
POEM	**:**	Patterns of Oriented Edge Magnitudes
RVFL	**:**	Random Vector Functional Link
ScSPM	**:**	State-of-art classification spatial pyramid matching
SLFNs	**:**	Single-Layer Feed-forward Networks

1. INTRODUCTION

The demand for higher level security needs has pushed the technology to expand to meet this demand. New technologies being developed are bound to have a certain feature to make them acceptable worldwide. They ought to be acceptable and less complicated to meet the needs of all the users globally. Therefore, the need for powerful, user-friendly schemes, capable of protecting the users' privacies and protecting their assets has attracted interests from diverse scholars. This field of study that concerns with the application of statistical analysis to biological data is called biometry. Biometrics refers to the measurement and statistical analysis of people's specific physical and behavioral characteristics. The underlying principle is that individuals can be identified by their physical and behavioral characteristics. There are diverse scientific notations of mathematical introductions for face recognition. These include areas such as pixel arithmetic for readers concerned with the mathematical representation and perspectives of a pixel in face recognition applications. The face recognition technology is based on mathematical principles. Therefore, researchers are indulged in efforts to find an algorithm that can be applied in solving all biometric systems. This implies that the face recognition technology is still unresolved issue that demand inputs from scholars around the world.

A quick search of the phrase Face Recognition in the Institute of Electrical and Electronic Engineers Journal reveals over 26,000 results with over 3,000 publications in journals and magazines. This is because the technology attracts interests from diverse fields as it is applicable in numerous areas. These include areas such as photography, virtual reality applications, human-machine interactions, and as well in the security applications. The diverse interests attract scholars from different fields in an attempt to solve some of the existing issues in the face recognition technology. Another reason for diverse interest is owing to the fact that the technology is also applicable in the recognition of the subjects and patterns, processing images, graphics, and as well applications in the social sciences such as psychology [1]. The topic of face recognition has been in existence for decades. The earlier focus was in the field of psychology through factors such as facial expression, perceptions, gestures, interceptions, and ability to read emotion in people in 1960s [2]. Bledsoe, alongside Bisson made some progress utilizing computers in the recognition of human faces [3-5]. There were some controversies on the source of funding

and therefore, some of the research papers were never published. Bledsoe developed and tested a semi-automatic application that was intelligent for the modern face recognition technology. A human operator selected some face coordinates, and then the system used this information for the recognition. The investigation identified a complex issue that has been a subject of debate for over 50 years. These include the variations in the illumination, expressions and ability to rotate the heads and aging among others in an effort to define unique features that can be used to identify an individual. To date, modern scholars still try to define the features such as the size of various facial features and organs to identify various individual features. An example was Lesk and Harmon [6] who applied the technology in Bell laboratories. They were able to define a vector that identified over 20 features such as the length of the nose, height of the ear protrusion, eyebrows as the foundation of the face recognition using the identification of such patterns. Their procedure was detailed and applied local template identical and as well the global measure of fit in the determination of parameters that would be used in the face recognition.

The 1970's was an interesting period as scholars sought to have a defining feature that will be applied in the future technologies. Some scholars opted for a definition of the face as a geometrical shape and then drew some precognitive patterns that would be applied in the recognition of the face. A major breakthrough came in 1973 when Kanade developed and applied a computer program that could carry out the task without the need for human intervention [7]. He developed a system that was applied in the recognition based on the extraction of 16 parameters without the need for human intervention. This was the first milestone towards facial recognition systems that did not need human intervention to implement. This system worked on the same principles as a human system as it would extract the needed parameters and identify the patterns in them in an effort to determine the main features. His accuracy rate was below 75% but was quick to note that the accuracy could be improved if some irrelevant features were to be left out. This was a foundation that guided the later studies in 1980's as subsequent scholars sought to build on Kanade's achievements. Other scholars tried different approaches such as template matching methods which were extrapolated with other strategies such as the application of deformable patterns. The decade also brought about new technologies such as the introduction of algorithms that employed the artificial neural networks. This was termed as eigenfaces and was the leading technology in the subsequent years.

Sirovich et al. were the creators of this new technology [8]. Their approach was founded on the principle of component analysis. Their efforts aimed at representing the image in a miniature without losing details. The saved dimensions could then be used in the reconstruction and these formed the bases for the subsequent studies that were carried out in the subsequent periods. While it applied by some scholars in the development of a procedure that was to be applied in the following decade. 1990 has brought a new phase in the technology as eigenface and further developments using the state of art technology and as well as the first industrial order was made. It obvious the new stage of commercialization of the technology that paved ways for further refinement and development. In another instance, Turk and Pentland of MIT made a breakthrough in advancing the eigenfaces technology of face recognition [9]. Their approach was unique because it was able to find a target head, save track and as well as classify it based on certain criteria. The technology has been further embraced and investigated even in the years after the 1990s. New strategies and procedures have been developed to improve on the previous models such as the Linear Discriminant Analysis, Principle component analysis, and the independent component analysis and as well their derivatives. This paper will elaborate more on other approaches and techniques that are applicable today. The paper will also highlight the evolution of the facial recognition technology and the milestones that have been achieved so far.

1.1. Face Recognition

Face recognition is one of the most effective apparatus that is used in the understanding and as well in the processing of images. In the recent past, face recognition technology has received a lot of attention owing to its diverse applicability. Some of the notable use security on photos and in videos as well. The technology has also been developed for use in such activities such as curbing election frauds. Numerous conferences have been organized in an effort to support the growth and development of the technology. An example is the Audio- and Video-Based biometric human authentication [10]. Another procedure is the unmanned face and expression recognition [11], empirical evaluations methodology of face recognition technology. Others include XM2VTS database [12] protocols and FERET. The growth in this technology has been fuelled by its applicability in the enhancement of the national security apparatus and as well a diverse commercial

application in other sectors. The technology has been in the development stage for over 30 years and has yielded a perfect technology that is effective and reliable. High demand for a simple but secure technology that is capable of protecting the privacy and assets without having to involve numerous statistical approaches has necessitated the development of the facial recognition technology. The modern technology uses personal identification numbers (PIN) to access services such as the automated teller machine (ATM), in others such as access to devices and computers, one need to key in a combination of passwords and numbers which requires physical involvement to gain access. There are other available technologies available today such as retinal or iris scans, fingerprint scans, and palm scans among others, they all rely on the client's cooperation and thus, the security personnel will have to install barriers and other restrictions to force people to use the security measures. Face recognition technology is ideal as it will not depend on the participants' as showed in Table 1.1. Willingness to collaborate and as well the information. The face recognition technology is therefore decent and more effective as compared to the others. The table below highlights some practical applications of the facial recognition technology.

Table 1.1. The typical practical applications of the technology

Areas	Specific Applications
Biometrics	Entitlement Programs ,Drivers' Licenses
	Passports ,Voter Registration ,Immigration, National ID
	Welfare Fraud
Information Security	Desktop Login (Windows, Windows NT,)
	File Encryption, Database Security, Application Security
	Access to internet services, Health Records, Intranet Security
	Secure banking and trading channels over the internet
Law Enforcement and Surveillance	Control of the CCTV and Video Surveillance, including application after an event.
	Investigating petty crimes such as shoplifting and as well as tracking suspects
Smart Cards	User Authentication in stores
Access Control	Vehicular Access and as well as buildings

The technology is applicable in many areas and one such application is in the identification of a face in a playing video or a captured scene. The technology is applied in the verification and picking of the face in a captured image or in a playing video. The system works in two ways; by comparing the face with the images in the database; and then releasing the required information such as the gender, and other identifying aspects of the target face. The application uses strategies such as segmentation of the faces from scrambled scenes and as well as narrowing some unique features on the face to verify or identify a face from the images. In this case, the input is a face that is not known and the system returns the results with the details of the unknown face in the identification case. On the other hand, in the verification process, the application rejects or accepts a claim that is made in regards to a particular face. The law enforcement and as well the commercial application of the face recognition technology can either be static, controlled environment, and uncontrolled images which can be static or dynamic. The nature of the images determines the nature of the application and as well the technique that will be used in processing the images. The images pose challenges that determine the quality of the obtained images and thus dictating the technique that will be applied in identifying the target faces. These include factors such as the image clutter and background noises affecting the ability to identify of clarifying a certain target face. The availability of a specific criterion that will be applied and as well the presence of separation algorithms determines the ability to get accurate outputs. In some cases, the objectives are to determine how a face would look like while ageing. Therefore, the algorithms, in that case, are only designed to transform an image and show the changes that would occur to an individual face over time. Therefore, depending on the desired outputs, the algorithms and the inputs are different to facilitate the creation of the targeted outputs.

The technology has undergone several transformation and developments to achieve its current state. In addition, it gives a substantial review of the procedures and tactics that were available by its idea. This preceded by another review in Chellappa et al. [15] that showed a huge transformation within a period of fewer than three years. A notable fact was that the facial recognition on dynamic targets such as the case in the videos was in the development stages. The video face recognition received massive interests from the stakeholders facilitating its growth and development to its current level. The process of face recognition is highlighted in the flowchart below. The process demonstrates the

phases of the procedure right from the input feeding, pre-processing, processing and the display of the output (Figure 1.1).
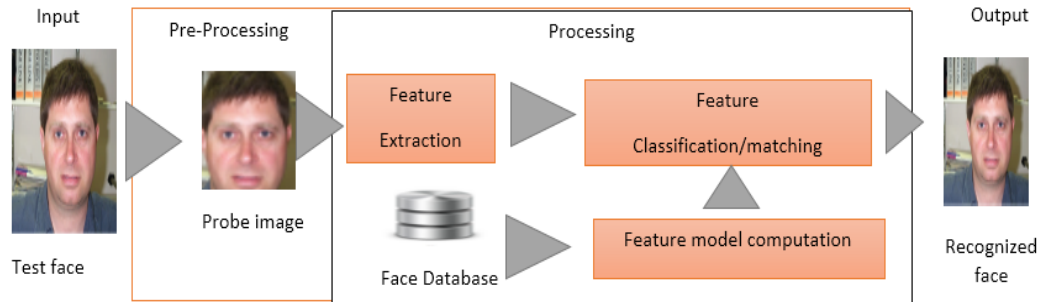
Figure 1.1. Face recognition Process

1.2. Literature review

In this chapter, we are going to provide the detailed survey of the face recognition investigations. There are literally thousands of papers has been dedicated to this evaluation by diverse scholars throughout the globe. There are two primary objectives of this study: the first is to offer some insights on the face recognition technology and the second is the provision an up-to-date review of the current publications on the topic. In an effort to offer and provide a comprehensive survey, existing face recognition procedures are detailed and categorized and representative techniques within each group are presented in the section below.

Wright et al. proposed multiple linear regression models to recognize the human face in frontal viewing by the sparse signal representation which it offers the key to addressing this problem [16]. Two algorithms have been applied in this investigation which is Sparse Representation-based Classification and Feature Extraction and Classification Methods, they used AR database which contains 4000 frontal images and some non-face images, the algorithm was able to reject nonface images, the result in this study was achieved 98% recognition rate.

Tan et al. proposed a face representation approach where authors used to combine the strengths of robust illumination normalization and local texture [17]. Authors alleviated the effects of changing illumination by employing a simple and efficient pre-processing chain. In addition, authors proposed Local Ternary Patterns (LTP) for describing the local

texture. Authors used Face Recognition Grand Challenge, Extended Yale-B, and CMU PIE datasets in their experimental works and obtained impressive results.

Zhao et al. proposed an OpenCV facial Haar-like structure, which is used to identify the face regions and the Principal Component Analysis (PCA) used for quick extraction of the face features and the Euclidean distance employed in the face recognition section [18]. The techniques showed that the test results showed that the system has a stable operation and high recognition rate.

Samaria et al. proposed a spatial observation classification method [19]. Authors extracted from a face image by using a group sampling method. Each face image was characterized by a 1D vector sequence of pixel observation. Each observation vector is a block of L lines and there is M lines overlap between successive observations. Unidentified face first tested to an observation sequence. Then it is matched in all HMMs face database. The match with the highest likelihood is considered the best match and the relevant model reveals the identity of the test face.

Hesher et al. realized principal component analysis (PCA) methods by using different numbers of the eigenvectors and the image sizes [20]. Each image has 6 different facial expressions in the face dataset for each of them. The performance reported result from using multiple images per subject in the gallery then it raised the recognition rate.

Yang et al. proposed a new procedure for computation of the covariance matrix using triple images where each image was taken in the different illumination condition [21]. Authors used the FERET database and the obtained accuracy was 95% for test set images.

Zhang et al. suggested a novel technique to extract illumination insensitive landscapes for face recognition under changing lights, which was called the Gradient-faces [22]. Gradient-faces method is robust to different illumination conditions. In addition, the gradient domain can discover the underlined structure of the images, for this propose they used PIE database which achieved 99.83% of the recognition rate and from the Yale-B database 98.96% recognition rates.

Liao et al. proposed a novel method for face image recognition which extracts texture features from face images [23]. The proposed method is robust against rotations of the face images. The method is less sensitive to histogram equalization. Authors compare two features, the first one is dominant local binary patterns and the second is supplementary features. Their result showed that this approach achieves better accuracy than other methods used in texture databases.

Sun et al. used the set of high-level approach representation as the deep hidden IDentity (DeepID) for the face confirmation [24]. The DeepID can effectively learn multi-classes through the challenging the faces. The DeepID increases more face classes at training section. Therefore this approach can add more than 10,000 faces in the training section and configured to keep reducing the number of neurons along the hierarchy in addition any state-of-art classifies can be learned based on their approach and they achieved 97.45% of recognition rates.

Bai et al. proposed an algorithm which can solve the problem of a very large number of training iterations on neural network (back-propagation) with the network calculating the weights in one iteration [25]. First, the dataset needs to be connected to the system, and then input sends the image to the hidden nodes. Before calculating the weight, the hidden node takes an output sample and sets it a weight according to the output nodes, therefore the nodes run with the best weights in just one iteration through the mapping between the hidden and output nodes as local receptive fields.

Chao-Kuei et al. demonstrated the vision and pattern recognition comprehensively brings out the aspect of face recognition in a robust manner [26]. In their study, the benefit of an unclear correspondence of feature point branding is integrated into a constrained optical flow algorithm that enables face recognition from a set of expressional face pictures. The information that is obtained through the calculation of self-optical flow is combined with the already integrated face image to carry out integrated face recognition in a probabilistic structure. This is a significant advancement that seeks to enhance the precision of face recognition in facial expressions.

Kabir et al. applied local directional number pattern (LDN) as a viable descriptor for facial analysis that utilizes the local features as well [27]. As seen earlier, LDN facilitates the encoding of the face's textures through directional information usage. An LDP operator to give a code with some strength computes the edge response values in all directions at each pixel location. Edge responses are preferred because they are more noise and illumination insensitive than intensity parameters in describing the local primitives.

1.3. Research Objectives

Face recognition correctness, by cover area information to minimize organization problems.

1. To enhance testing accuracy using Extreme learning method based –LRF approach with the various dataset.
2. To improve our expert system as compare to existing author literature study.
3. To calculate the average testing accuracy for diverse noisy environment datasets.

1.4. Thesis Statement and Organization

This thesis is categorized into four sections as detailed below:

Section 1 – Introduction: It highlights the thesis statement, problem statement and overview, background research on the face recognition and as well, the history of the face recognition technology and literature review about face recognition applications.

Section 2- We are going to describe face recognition algorithms such as neural network and extreme learning machine as well.

Section 3– Result: This is dedicated to the evaluation and as well the assessment of the outcomes of the new approach.

Section 4 – Conclusion: This chapter will harmonize the thesis, research objectives, plus giving inference to the conclusion.

2. RELATED THEORIES

2.1. Human's Brain

The knowledge concerning how a human mind works perfectly has remained unknown over time. However, all they understand are the set of information are processed. The function of neurons is to gather signals emanating from other neurons through dendrites. Once the neurons receive the signals, they are transmitted through axon in the form of electrical action (Figure 2.1). The axon further divides into millions of branches at its end to form a structure known as a synapse. The role of a synapse is to convert action or the signal received via axon to an electrical effect, how those neurons convey information or electrical effects from one to another. What is the nexus between the neurons as far as the transfer of signals is concerns. In a situation where the neurons are equipped with enormous input relative to inhibitory inputs, electrical action spikes are usually sent directly to the axon. The learning usually takes place whenever information is transmitted across the neurons through synapses. Therefore, the changes inherent between neurons facilitates learning (Figure 2.1).
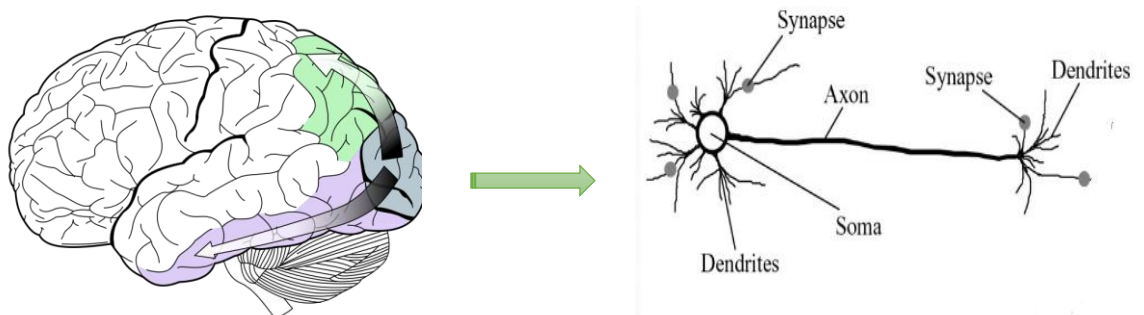

Figure 2.1. Human's brain structure

2.2. Artificial Neuron

There exist different kinds of artificial neuron that affect the ability of man to think through electrical replication of the human's thinking procedure. It should be remarked that normal neurons could be subdivided into four major categories, which include soma,

axons, dendrites, and synapse. The process of categorizing those neurons takes into account their biological titles. The role of dendrites in a neuron is to receive stimuli or input network obtained from the soma. The input network is then conveyed to other neurons via synapses. Eventually, the soma absorbs and interprets the input network to provide an input, which is transmitted to other neurons via axon and synapses. According to the results obtained from recent research, from a mechanical approach, biological neurons are usually more complicated in comparison to the little explanation provided below. They are considered more involved relative to the contemporary artificial networks, which is a development of the modern artificial neural networks. Since biology can denote seen as a key player in improving human neurons and ensuring equipment progression, net designers have the responsibility of designing and coming up with more information about the human brain. However, it should remain perceived that artificial neural network is less concerned with superlative regeneration of human mind. Contrary, neural network scholars have initiated a research geared towards comprehending the capacity of nature through which people can get solutions for the issues that have not been solved by the ancient computing. To facilitate the process mentioned above, neural network takes into account the four primary purposes of natural neurons. Figure 2.2 is a clear exhibition of an essential illustration of a neural network.
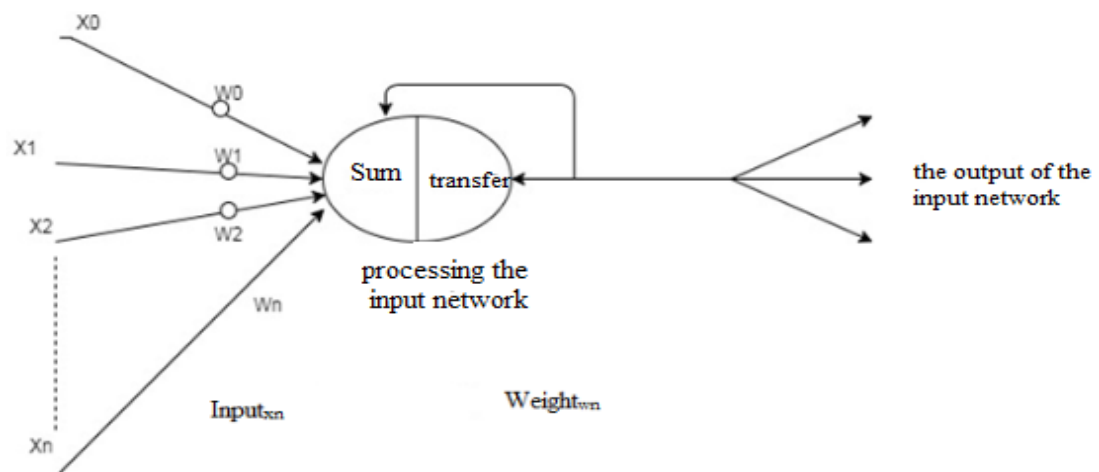
Figure 2.2. Basic Artificial Neuron

2.3. Neural Networks

Neural networks can be described as a system which as an interconnection of some processing elements whose operation is analogous and whose function is directed by the structure of the network, strength of the connection, and the process that takes place during the element or nodes computing. The inherent biological systems entirely influence the architecture of the neural network. Such systems are usually applied modest processing elements with the aim of getting high computation rates. The neural network symbolizes the brain majorly in two ways; the learning process is a network through which knowledge is obtained while the synaptic weights is an interneuron connection through which information and knowledge are recorded and saved. Practically, this translates to that fact that it is possible to construct a machine which can think intelligently [28].

Despite the fact that, it is possible to construct machines and install artificial intelligence, humans are still unaware on how the human brain can perform multiple functions such as calculations, thinking, and operations simultaneously, even if the human mind is faster and has a smaller memory relative to machines. It is unknown on how the brain is manipulated and condition to perform hardware parallelism. The elements of computing are usually organized in a way that millions of them handle the same problem simultaneously. The presence of the millions of neurons makes it easy to neutralize the skewness of computing powers where elements which are slow are combined to obtain significant results.

A neural network can further be recognized as an assembly of interrelated processing units, elements or nodes whose purpose is depended on an animal neuron. The network's ability to process is usually stockpiled in the weights commonly referred to as inter-unit connection strengths. Such weights are typically acquired through learning from, or adaptation to particular training patterns. To have a clear picture of the inherent processing differences, it is advisable to consider investigating the approach used by conventional computers inclusive of machines.

Face recognition using neural network algorithms works by taking one or more networks and applying them directly to image portions and then arbitrate the outcome. Each network is eventually conditioned to provide results either in absence or presence of a face. The process of conditioning a neural network to recognize a face is usually faced with

the challenge of categorizing the non-face images. More information about the appropriate structure of those images will be analyzed and discussed in the chapter for data set.

2.3.1. How Neural Network Works?

Artificial neural networks can change their behavior to respond to their surroundings. For instance, if they are provided with individual inputs, which probably have a predetermined output, they modify themselves to provide consistent responses. This aspect has earned them much attention. As such, different training algorithms have been established (Figure 2.3).
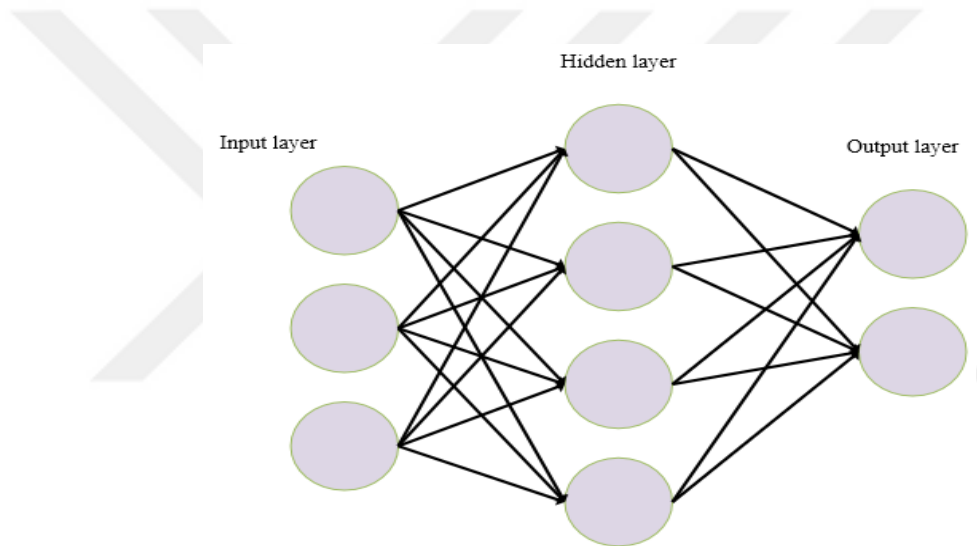
Figure 2.3. Neural network connection between (input layer, hidden layer, and output layer)

2.3.2. Multi-layer neural network

The multi-layer neural network is composed of numerous neuron layers. A neuron in one layer is directly connected to neurons in the subsequent layer and can learn sophisticated and nonlinear functions such as the XOR-function (Figure 2.4).
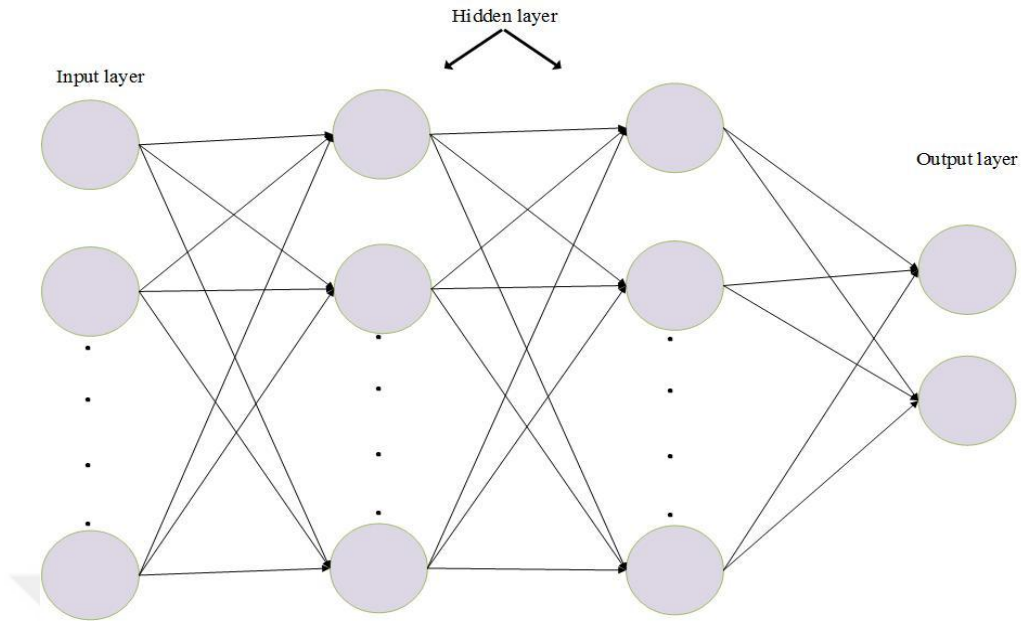
Figure 2.4. Multi-layer perceptron diagram

$$S = w_1 x_1 + w_2 x_2 + \theta \qquad (2.1)$$

The output of the perceptron is zero when S is negative and equal to one when s is positive. For a constant θ, the output of the perceptron is similar to one on one side of the dividing line which is defined by;

$$w_1 x_1 + w_2 x_2 = -\theta \qquad (2.2)$$

Figure 2.5. XOR-Problems in Single Layer Network

The multi-layer neural network can provide a solution to XOR-Problem. However, this is impossible during its complex structure that cannot be trained easily like the single-layer neural network. The issue of the multi-layer neural network training had remained for quite some time until a time in which back-propagation was developed. A single scientist thought an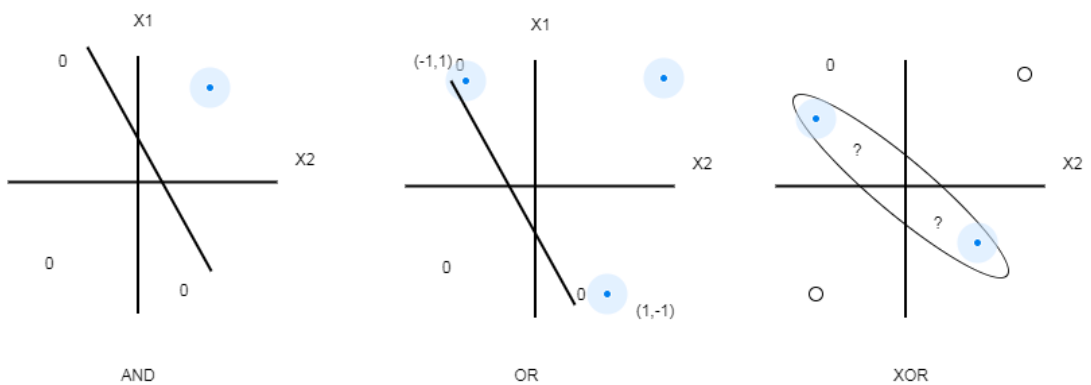 idea and developed it to form a new algorithm (Figure 2.5)., which was back-propagation. More information concerning how the back-propagation can be used to train the multi-layer neural network will be discussed in the next section.

2.3.3. Back-Propagation

According to previous chapters, the multi-layer networks can compute a broader range of Boolean functions in comparison to other networks that have single-layer computing units. However, the computational energy required for finding the appropriate combination of weights proliferates significantly especially when more parameters inclusive of sophisticated topologies are taken into account.

The objective of the chapter is to elucidate the general learning approach, which can handle extreme learning problems (the back-propagation algorithm). The back-propagation algorithm is a numerical method that is applied by various research communities in a variety of contexts. The method discovered and rediscovered over time. However, it was until the year 1985 when the connectionist artificial intelligent recognized the presence of the backpropagation algorithm via the work provided by the personal digital assistant group. Since then, the method has been used many times and researched techniques for neural networks learning. The method boasts of the merits such as more general relative to other analytical derivations and also easier to understand. Also, the backpropagation algorithm provides clear proof of how an algorithm can be efficiently implemented in computing systems and in the context where only the local data can be transmitted via a network.

The BP algorithm utilizes the gradient descent technique to search for the less or minimum error function primarily in weight space. The formation of groups which reduces the error function is determined to provide an elucidation to a learning problem. Although this approach involves calculations for gradient meant to give the error function at every stage, there is a mandatory assertion for the differentiability and continuity of the inherent error functions or purposes. Apparently, we have to make use of an activation function

rather using perceptron step function because a composite function formed by the interconnected perceptron is usually spasmodic and the hence the error. Sigmoid, whose real equation is SC: IR → (0, 1), is recognized as activation functions utilized by BP networks. The expression provided below is the definition of Sigmoid.

$$s_c(x) = \frac{1}{1 + e^{-cx}} \tag{2.3}$$

The constant 'c' is obtained arbitrarily, and its reciprocal $(1/c)$ is referred to as the temperature parameter in the contexts of stochastic neural networks. According to Figure 2.6., the sigmoid shape changes concerning the changes in the value of c. The graph illustrates the sigmoid for $c = 1, c = 2$, and $c = 3$. To solve all the expressions obtained from this chapter; it will be assumed that $c = 1$. However, after reading and contextualizing the materials an individual should be capable of simplify all expressions for c (sigmoid will be $sl(x)$ to $s(x)$)).
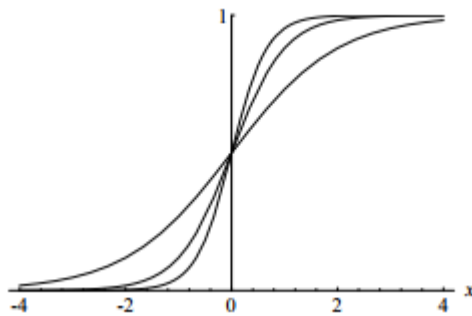
**Figure 2.6.** Three sigmoid (for c = 1, c = 2 and c = 3).

The derivative of the sigmoid with respect to x is

$$\frac{d}{dx}s(x) = \frac{e^{-x}}{(1 + e^{-x})^2} = s(x)(1 - s(x)). \tag{2.4}$$

The equation above represents the hyperbolic tangent for the argument x/2 and its shape has been provided in Figure 2.6. According to the figure, there are four kinds of squashing functions which are continuous in nature.

$$s(x) = 2s(x) - 1 = \frac{1 - e^{-x}}{1 + e^{-x}} \qquad (2.5)$$

This is nothing but the hyperbolic tangent for the argument x/2 whose shape is shown in Figure 2.7 (upper right). The figure shows four types of continuous "squashing" functions. The ramp function (lower right) can be used in learning algorithms taking care to avoid the two points where the derivative is undefined.
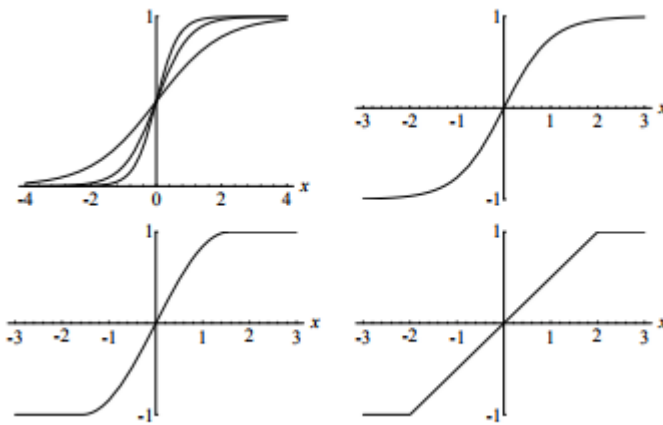
Figure 2.7. Graphics of some squashing function [40].

Recommendations for numerous activation functions have been provided through which the backpropagation algorithm can be implemented to them. A differentiable activation function can render the computed function differentiable because the network takes into account the function compositions. Also, the error function is also differentiable because of the gradient. The derivative of a sigmoid is always positive just like the gradient (error function).

2.3.4. Local minima of the error function

A value has to be rewarded for all the positive landscapes of the sigmoid as the activation function. The most important problem is that local minima appear in the error function which would not be there if the step function had been used. Figure 2.8 demonstrates an example of a local minimum with advanced error level than the other in the regions. The function was computed for a single unit with two weights, constant

threshold, and four input-output patterns in the training set. There is a valley in the error function and if gradient descent is started there the algorithm will not converge to the global minimum.
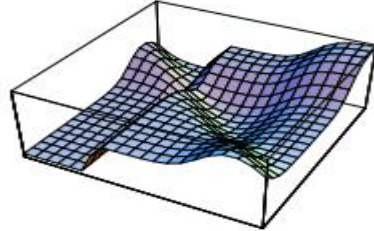
Figure 2.8. Local minima of error function [40].

In many cases, local minima appear because the targets for the outputs of the computing units are values other than 0 or 1. If a network for the computation of XOR were trained to produce 0.9 at the inputs (0,1) and (1,0) then the surface of the error function develops some protuberances, where local minima can rise (Figure 2.8). In the case of binary target values, some local minima are also present, as shown by Lisboa and Perantonis whom they are analytically found all local minima of the XOR function [29].

2.4. Convolution Neural Network (CNN)

The limitation of face recognition from 2D images is naturally not well posed for example there are different ways, which describes the training points well however they do not simplify well to hidden images. In a simplified manner is that they did not establish enough spaces for training points by the inputs images in order allow correct estimation of class probabilities through the input space. However, there is no invariance to local deformation or translation of the images for multi-layer perceptron networks with 2D images as input. The Convolution network uses three thoughts in achieving and constraining some degree of shift and distortion invariance (Figure 2.9).

Using shared weights also decreases the number of parameters in the system helping generalization. Set of layers contained in a typical convolutional network usually each of them covers one or more planes. Approximately normalized and centered images move to the input layer and resolve there. Every component in a plane receives input generated from a small neighborhood in planes of the layer before or layer earlier. The connection units to LRF idea dates back to the early 1960s during the finding of locally sensitive,

orientation-selective neurons in the cat's visual system by Hubel and Wiesel's [30]. Here the starting weights of the receptive field for a plane are forced to be the same at all points across the plane. Carefully each plane can be presented as feature map with a fixed detector feature that is convolved local spaces scanned over the planes of the layers before. For many features to be noticed there will be needed to use typically several layers. These layers are known as convolutional layers. The exact location of a feature has less significant if the feature has been sensed, therefore, the convolution layers usually overly over each other which makes a local subsampling operation and averaging for example as shown below.

The usual backpropagation gradient-descent process [31] helps in the training of the network. To reduce the number of weights in the network, connection strategy can be used which is done by connecting the feature maps in the second convolutional layer unique to 1 or 2 of the maps located in the first subsampling layer(the connection strategy was chosen physically).
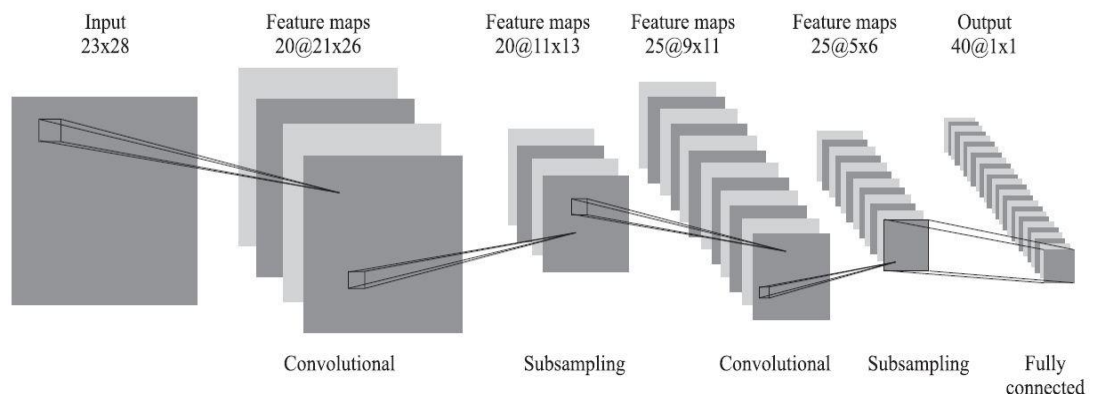
Figure 2.9. The representative of convolutional network [43]

2.5 Extreme learning machine overview

Extreme learning machine is one of the major variants of the artificial neural network with single hidden layer Feed Forward architecture (SLFNs). This method has been developed to overcome the weakness of Artificial Neural specifically on the process of learning based on the reasons described by Huang et al [32]. Artificial neural network using Gradient Descent learning method is used in in the learning process where there are a long process and too many parameters to be determined.

The following procedure shows the steps of Mathematical Model of the extreme learning machine as follows;

Given N sample data (x_i, t_i), where

$$x_i = [x_i1, x_i2, \ldots, x_in] \text{ then } T_i = [T_i1, T_i2, \ldots, T_in] \qquad (2.6)$$

And Ñ is the sum of Hidden neurons and $g(x)$ is an activation function that can be modeled as follows:

$$\sum_{i=1}^{\tilde{N}} \beta_{ig} (W_i . X_j + b_i) = t_j, j = 1, \ldots, N. \qquad (2.7)$$

Given $w_i = [w_{i1}, w_{i2} \ldots w_{in}]$ is the weighted value connected between the input layer and the hidden layer, and $\beta_i = [\beta_{i1}, \beta_{i2} \ldots \beta_{in}]$ is the value of the weights that are connected between the hidden layer and the output layer. The bi-value is the bias value of the ith node. w_i, x_j is the inner product on w_i and x_i. The activation function g (x) can approach N data samples with zero error values can be modeled as follows:

$$\sum_{j=1}^{\tilde{N}} ||oj - tj|| = 0 \qquad (2.8)$$

The above equation can be written as follows;

$$T = H\beta$$

Where H, β and T are described as follows;

$$H = (w_1, \cdots, w_{\tilde{N}}, b_1, \cdots b_{\tilde{N}}, x_1 \cdots, x_{\tilde{N}})$$

$$= \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \cdots & g(w_{\tilde{N}} \cdot x_1 + b_{\tilde{N}}) \\ \vdots & \cdots & \vdots \\ g(w_1 \cdot x_N + b_1) & \cdots & g(w_{\tilde{N}} \cdot x_{\tilde{N}} + b_{\tilde{N}}) \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix} [\tilde{N} \times m] \qquad (2.9)$$

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_{\tilde{N}}^T \end{bmatrix}$$

H is the matrix of the hidden layer, the ith column is the result of the hidden neuron to ith based on the input $x_1, x_2, \cdots, x_n$. To perform the train in process with minimum error can use the equation below:

$$\|H\beta - T\| = \|HH^{\dagger}T - T\| = \frac{min}{\beta}\|H\beta - T\| \tag{2.10}$$

$$\beta = H^{\dagger}T$$

Where $H^{\dagger}$ is the Moore-Penrose generalized inverse matrix of H.

In short, the ELM method can be done in several steps:

Step 1: Randomly initialize w_i and bi values

Step 2: Calculate the result of the hidden layer matrix H

Step 3: Calculate the weight of the value of β

2.5.1. Local receptive fields based on extreme learning machine (ELM-LRF)

First, this machine was established for Single-Layer-Feed-Forward Neural networks (SLFNs). Afterwards, it stretched out the concept to include concealed neurons. Relatively unique to the normal comprehension of neural systems, ELM hypotheses [44]– [46] demonstrate that although hidden neurons assume basic parts, they require not be balanced. Neural systems can accomplish learning ability without iteratively tuning hidden neurons as long as the initiation elements of these hidden neurons are nonlinear hybrid function ceaseless including however not restricted to organic neurons. The organic neurons yield capacity can not be unequivocally communicated with solid formula since distinctive neurons may involve diverse yield capacities.

The yield elements for natural neurons are a nonlinear hybrid function which has been considered by the hypothesis of ELM. The Unsystematic Vector Functional-Link (RVFL) [33] and Quick Net [34] utilize direct connections for the input and outputs. It utilizes a particular ELM shrouded layer called the sigmoid. For deep examination, scholars may allude the relationship by contrasting the ELM amongst RVFL/Quick Net) [35]. ELM initially changes the input by ELM mapping feature to a shrouded layer. At that

point, the outcome is produced by learning the ELM. The learning involves characterization and ELM highlight mapping,

The ELM yield capacity for summed up SLFNs is given by:

$$f(x) = \sum_{i=1}^{l} \beta i h_i(x) = h(x)\beta \tag{2.11}$$

Where $\beta = [\beta \dots \beta]T\ g$ and b is the vector of the output weights between the hidden layer with L hubs to the yield layer with $m \geq 1$ hubs

$h(x) = [h1(x\dots.h1(x)]$ is the output (row) vector of the hidden layer. Distinctive initiation capacities might be utilized as a part of various hidden neurons. Specifically, in genuine applications hi(x) can be

$$h_i(x) = g(a_i, b_i, x_i)a_i \in R, b_i \in R \tag{2.12}$$

Where $G(a, b, x,)$ is a nonlinear piecewise continuous function and (ai, bi) are the $i - th$ hidden node parameters. Some generally utilized actuation capacities are:

Sigmoid capacity:

$$G(a, b, x) = \frac{1}{1 + \exp(-(a.x + b))} \tag{2.13}$$

Fourier function

$$G(a, b, x) = \sin(a.x + b) \tag{2.14}$$

Hard limit work

$$G(a, b, x) = \left\{ \begin{smallmatrix} 1\ if\ a.x - b \geq 0 \\ 0, otherwise \end{smallmatrix} \right\} \tag{2.15}$$

Gaussian capacity

$$G(a, b, x) = \exp(-b\|x - a\|^2 \tag{2.16}$$

Multi-quadrics work

$$G(a, b, x) = (\|x - a\|^2 + b^2)^{1/2} \tag{2.17}$$

Every hidden node of ELM G (a, b, x) can even be a supernode consisting of a sub-network. This is not the same as stipulated by Schmidt et al., and RVFL model of which they only use sigmoid or RBF hidden nodes type. The universal estimation and grouping ability of Schmidt, et al., model have not been demonstrated before ELM theory are proposed its general estimate capacity will wind up plainly evident when ELM theory is linked. Igelnik and Pao [46] demonstrates that RVFL's universal estimation capacity when semi-irregular hidden node is utilized, that is, the input weights (a_i) are randomly created while the hidden node biases (b_i) are computed based on the training samples (x_i) and the input weights (a_i)

In contrast to RVFL theories for semi-randomness, ELM theories demonstrate that all the hidden nod parameters can be randomly produced as long as the initiation function is nonlinear piecewise continuous. Similarly, all the hidden nodes can be autonomous from preparing tests as well as free from each other. ELM theories are substantial in sigmoid, RBF threshold, and trigonometric networks etc.

2.5.2. Local Receptive Fields

Here local receptive fields based on (LRF) are proposed as a non-specific ELM engineering to solve image processing and comparable issues in which diverse density of connections are requested. These connections between input and hidden nodes are inadequate and limited by comparing open fields, which are tested by any nonstop likelihood appropriation. Also, combinatorial nodes are utilized and this results into a translational invariance to the network by consolidating a few hidden nodes together. It involves no gradient-drop steps and the training is astoundingly effective.

2.5.3. Hidden Nodes in Full and Local Connections

ELM theories demonstrated that hidden nodes could be created randomly as per any likelihood distribution. In the real case, the randomness is two-phases:

1. Different density of connection amongst input and hidden nodes are randomly sampled because of various sorts of likelihood conveyances utilized as a part of utilization. The weights amongst input and hidden nodes are additionally produced. Completely linked hidden nodes, as appeared in Figure 2.10, have been broadly examined. Built from them,

ELM accomplishes best in class execution for many utilizations, for example, remote detecting, time arrangement examination, content order [37], activity acknowledgment, and so forth. Therefore, these works concentrate just on irregular weights, while disregarding the characteristic of arbitrary links.

For common images and dialects, the solid neighborhood relationships may make the full links less fitting. To address the issue, we propose to utilize hidden nodes which are privately linked with the inputs. As per ELM theory, the links are arbitrarily inspected in light of certain likelihood circulations, which are denser around some information hubs while sparser more distant away

2. Local Receptive Fields Based ELM: a locally strong connection case of ELM-LRF is given in Figure 2.11 in which the links between the input layer and hidden node (i) are randomly created by some constant probability distribution. Some receptive field is produced by this random association. The links are strong around some input nodes and sparse farther away (yet do not completely disappear in this example). Strong natural proof additionally legitimizes the local receptive fields, which demonstrates that every cell in the visual cortex is just sensitive to a corresponding area of the retina (input layer). ELM-LRF learns the local structures and produces more important portrayals at the hidden layer when managing image processing and comparable errands.

3. Combinatorial Node Local receptive fields may have diverse structures in natural learning, and they might be created in various ways utilizing machine-learning methods. One of the techniques is the combinatorial node recommended in ELM theories. ELM demonstrates that a hidden node in ELM could be a blend of a few hidden nodes or a sub-system of hubs. For instance, local open fields in Figure 2.12 may be considered as a situation where the combinatorial node is framed by a sub-network (as grouped by the circle). The output of this sub-network is a summation (which require not be direct) of the output of three hidden nodes connecting to corresponding local receptive fields of the input layer. Therefore, the element created at one area (one specific hidden node) has a tendency to be valuable in various areas.
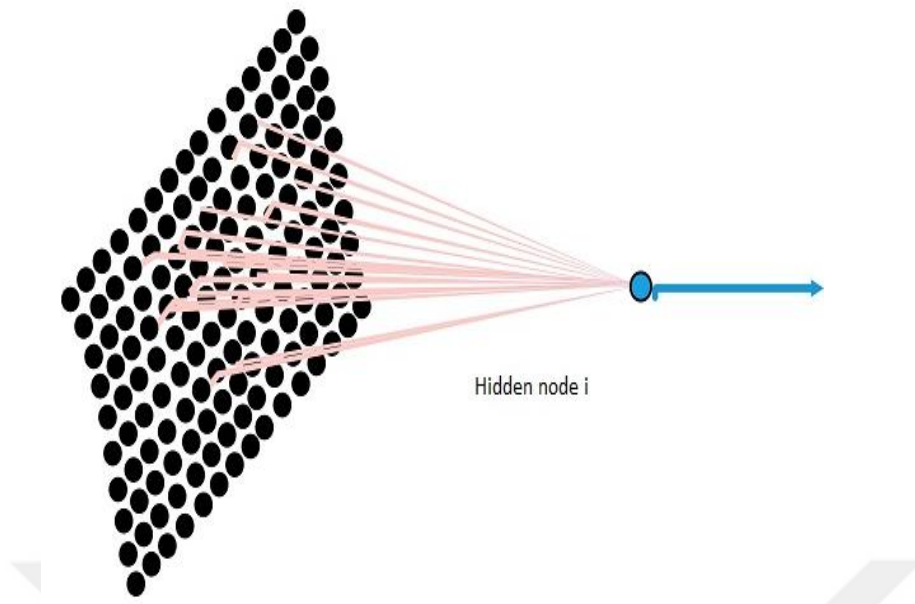
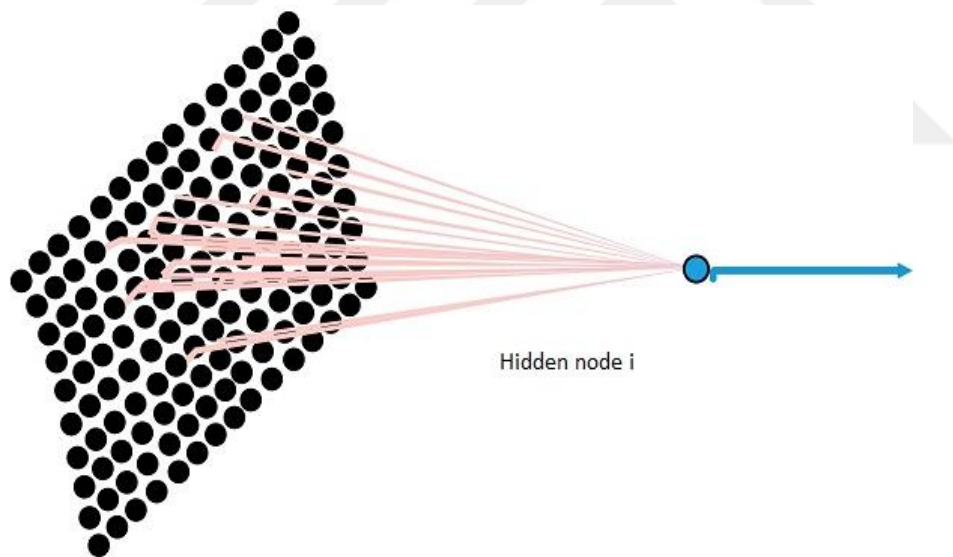**Figure 2.10.** Extreme learning machine hidden node in full connection

Figure 2.11. ELM hidden node with local connection
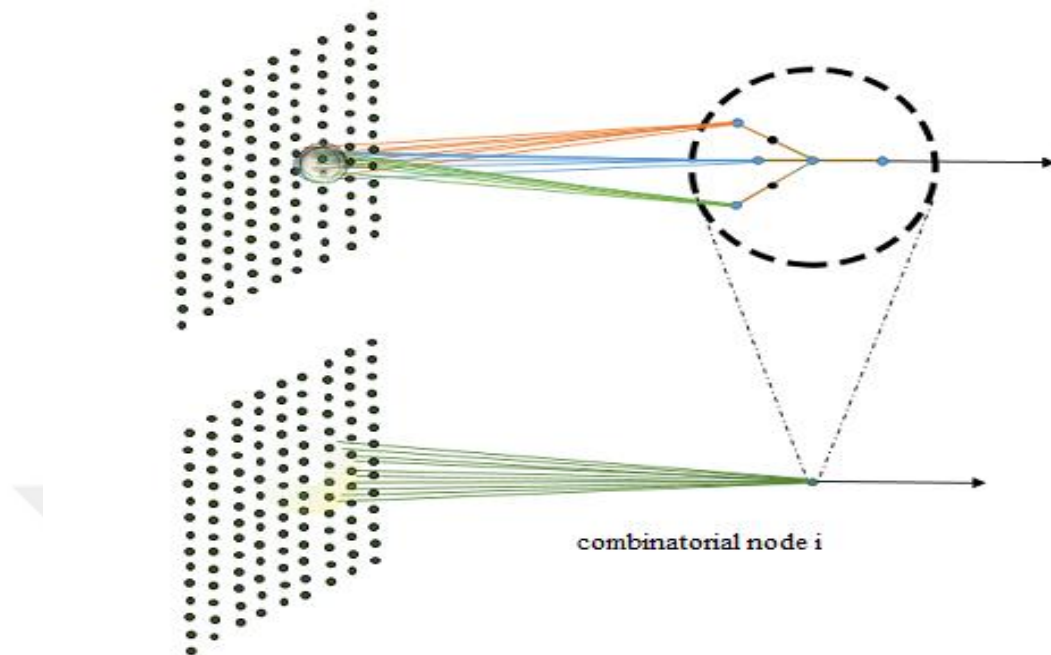
Figure 2.12. The combinatorial node of ELM: a hidden node can be a sub-network of several nodes, which turn out to form local receptive fields and (linear or nonlinear) pooling.

Accordingly, the ELM-LRF system will be invariant to minor interpretations and pivots. Also, the associations between the info and combinatorial nodes may take in the local structures far superior. Denser connections around the information hub due to the cover of three open fields (see Figure. 2.12) while sparser more remote away. Averaging and max-pooling techniques are clear to figure these combinatorial hubs. Moreover, the learning-based technique is likewise potential to develop them.

2.5.4. Specific Combinatorial Node of ELM-LRF

Although diverse sorts of nearby responsive fields and combinatorial hubs can be utilized as a part of ELM, for the accommodation of execution, we utilize the basic advance likelihood work as the inspecting conveyance and the square/square-root pooling structure to frame the combinatorial hub. The responsive field of each shrouded hub will be made out of info hubs inside a foreordained separation to the middle. Besides, just sharing the info weights to various shrouded hubs straightforwardly prompts the convolution operation and can be effectively actualized.

We can construct a particular case for the general ELM-LRF:

1. Random convolutional hub in the element outline a sort of the privately associated shrouded hub

2. A node in the pooling map in Figure 2.12 is a case of the combinatorial hub.

2.5.5. Random Input Weights

To obtain through representation of the input *K*, distinctive input weights will be used to produce K diverse feature maps. In Figure 2.13, hidden layer is made out of irregular convolutional hubs. In addition, the input weights of a similar component maps are shared while distinct among various maps. The input weights are arbitrarily produced and after that are orthogonalized:

1. Create the underlying weight framework $\bar{A}_{init}$ randomly. With the input estimate d*x and receptive field r*r, the size of the feature map ought to be $(d - r + 1)(d - r + 1)$.

$$\bar{A}^{init} \in R^{r^2 \, x \, k}$$

$$\hat{a}_k{}^{init} \in R^{r^2} \, k = 1, \dots K$$

(2.18)

Orthogonalize the underlying weight framework $\bar{A}^{init}$ utilizing particular esteem decay (SVD) strategy. Eq 2.18 the orthonormal premise of the orthogonalization enables the system to remove a more complete set of features than non-orthogonal ones. Subsequently, the speculation performance of the network is additionally improved. Orthogonal random weights have additionally been utilizing and then create incredible outcomes.

The input weight to the $R - th$ feature map.

$$C_{i,j,k}(x) = \sum_{m-1}^{r} \sum_{n-1}^{r} x_{i+m-1,j+n-1} \cdot a_{m,n,k}$$

$$i, J = 1, \dots, (d - r + 1)$$

(2.19)

A few works have exhibited that CNN with certain structure likewise can perform well with ELM feature mapping (through nonlinear mapping after arbitrary info weights. In this paper, the proposed ELM-LRF, where the input weights to the feature maps are

orthogonally random can accomplish far and away superior execution than all around trained counterpart (Figure 2.13).
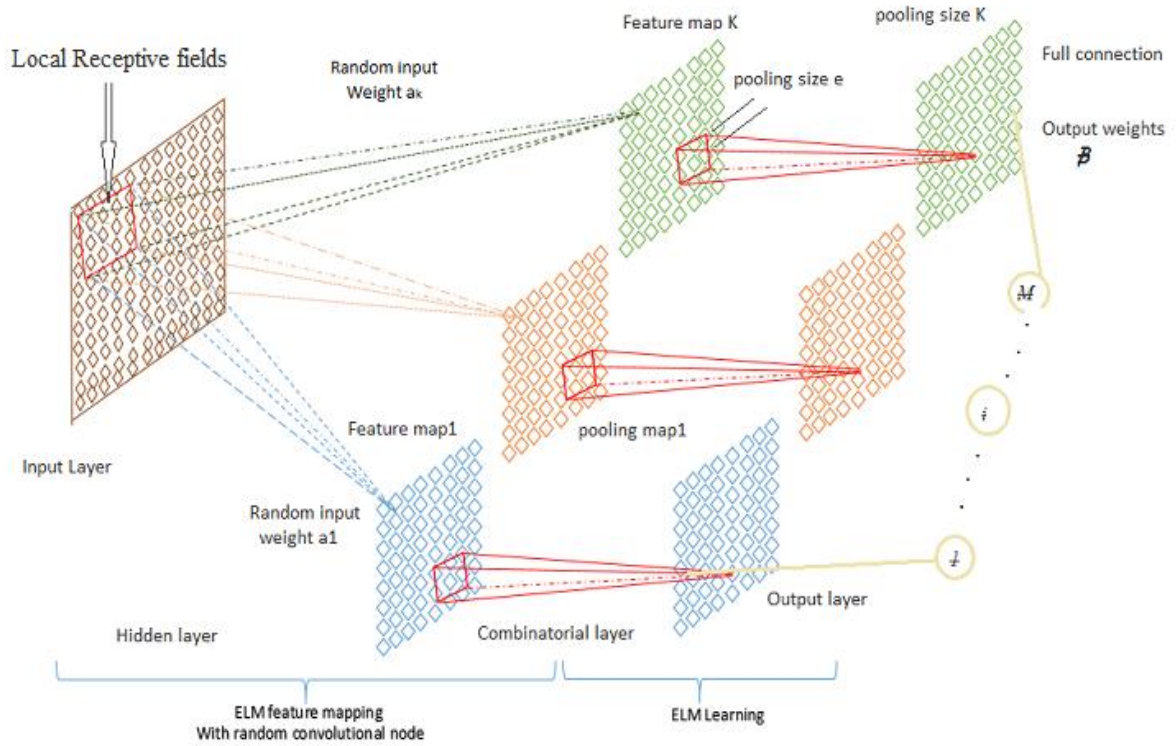
Figure 2.13. The implementation network of ELM-LRF with K maps.

3. EXPERIMENTAL WORKS AND RESULTS

Extensive experimental works have been carried out to show the efficiency of the proposed ELM-LRF method on face recognition. All experimental works were designed and run on MATLAB environment on a computer having an Intel Core i7-4810 CPU and 32 GB memory. Three face recognition datasets namely; UFI Face dataset, Caltech face dataset and CBCL Face dataset, respectively. The UFI dataset consists of 478 training images and 478 images for testing, with the image size cropped to 32×32 pixels. The second dataset applied was the Caltech face dataset, cropped 64×64 pixels. The MIT CBCL dataset contains 2,429 training set faces and 4,548 non-faces, in addition to 472 testing faces and 23,573 non-faces. These datasets are publicly available for academic research works. The details about the datasets can be found in the next sub-sections.

The ELM-LRF structure that we have used in our experiments had 4 layers. These layers are input layer, convolution layer, pooling layer and ELM classification layer, respectively. The input layer is responsible to get the resized input images and convey them into the convolution layer. This layer also covers several pre-processing steps such as normalization. The convolution layer contains adjustable filters. The number of filters and their coefficients are assigned randomly. Pooling operation is employed after filtering operation. To this end, max-pooling or average pooling can be applied. In our experiments, the convolution layer contains various numbers of filters of various. The obtained results are evaluated with classification accuracy, which is defined as the ratio of the number of correctly classified samples to a number of all samples.

As it was mentioned earlier, three face datasets have been used in experiments and the obtained results are recorded in the related Tables. We further compare the obtained results with several methods such as Local Gabor Binary Pattern Histogram Sequences (LGBPHS), Local Derivative Pattern Histogram Sequences (LDP-HS), Face Specific Local Binary Pattern (FS-LBP) and Patterns of Oriented Edge Magnitudes Histogram Sequences (POEMHS) [39]. In LGBPHS method, a face image is modeled as a histogram sequence by concatenating the histograms of all the local regions of all the local Gabor magnitude binary pattern maps. For recognition, histogram intersection is used to measure the similarity of different LGBPHSs and the nearest neighbourhood is exploited for final classification. LDP extracts high-order local information by encoding various distinctive

spatial relationships contained in a given local region. Then similar to LBP, a histogram sequence can be calculated from LDP coded images. POEMHS is an oriented spatial multi-resolution descriptor capturing rich information about the input face images. It is a multi-scale self-similarity based structure that results in robustness to exterior variations. It is of low complexity and practical for real-time applications.

3.1. Unconstrained Facial Images (UFI) Dataset

UFI database covers a series of real face images chosen from a large set of photos owned by the Czech News Agency [40]. Each face image is annotated with the name of the person and some images have some background objects. The dataset is composed of two partitions. The first part is called cropped images, which contains 605 automatically detected faces images. Images in the second partition, called large images, contain not only 530 face images, but also some background objects are presented in the images. Figure 3.1 shows some representative face image samples for UFI dataset.

Figure 3.1. UFI face dataset

As it was mentioned earlier, 478 training images and 478 testing images from UFI dataset was considered in experimental works. The cropped image sizes are 32×32 pixels. A four layered ELM-LRF structure was employed for face recognition task. The convolution layer contains 40 filters of size 3×3. The max-pooling is adopted in pooling layer. In Figure 3.2, the obtained face images after convolutional and pooling layer are given.

(a)   (b)

Figure 3.2. (a) Obtained face images after convolutional layer, (b) Obtained face images after pooling
layer

The obtained results are tabulated in Table 3.1. As seen in Table 3.1, the best result is obtained with POEMHS method where the accuracy is 67.11%. The second highest result is obtained with ELM-LRF method, where 66.11% accuracy is recorded. The worst result is obtained with LDPHS method, where the accuracy is 50.25%.

Table 3.1. Results for the UFI dataset

Method	Testing Accuracy (%)
LBPHS	55.04
LDPHS	50.25
POEMHS	67.11
FS-LBP	63.31
ELM-LRF	66.11

3.2. Caltech Face Dataset

The Caltech face image dataset was constructed from various face images that were collected from the websites. The datasets also contain the coordinates of the eyes, the nose and the center of the mouth for each frontal face are provided in a ground truth file. The ground-truth information can be used to align and crop the human faces for a face

detection algorithm. 10,524 face images were collected which have various resolutions. Various images for two people are given in Figure 3.3.

Figure 3.3. Caltech face dataset

From Caltech face image dataset, we cropped 7366 face images of 64×64 pixels for training of the ELM-LRF and the rest face images were used for the test. In the convolution layer of the ELM-LRF, 30 convolutional filters of size 5×5 were considered. The average pooling was also considered in pooling layer. The obtained results are tabulated in Table 3.2. As seen in Table 3.2, ELM-LRF obtained the highest accuracy where the accuracy is 98.15%. The second best accuracy is recorded with SPARSE BASED NN method where its accuracy is 92.91%. The worse accuracy is obtained with NR MODEL where 32.36% accuracy is recorded.

Table 3.2. Results for the Caltech dataset

Method	Testing Accuracy (%)
NR MODEL	32.36
SCSPM	82.83
TSR	37.45
SPARSE BASED NN	92.91
ELM-LRF	98.15

3.3. MIT CBCL Face Dataset

The MIT CBCL face image dataset contains totally 25573 images of size 19×19. All images are in grayscale level and PGM format. This dataset just designed to detect an

image as a face or non-face. To this end, in the training set, 2429 face images and 4548 non-face images were considered. Besides, 472 face images and 23573 non-face images were in the test set of the dataset. In Figure 3.4, we show some examples face images from CBCL dataset. All images were resized to 32×32 before running the proposed ELM-LRF method. 40 filters of size 3×3 were used in the convolution layer and the max pooling operation was handled in pooling layer.
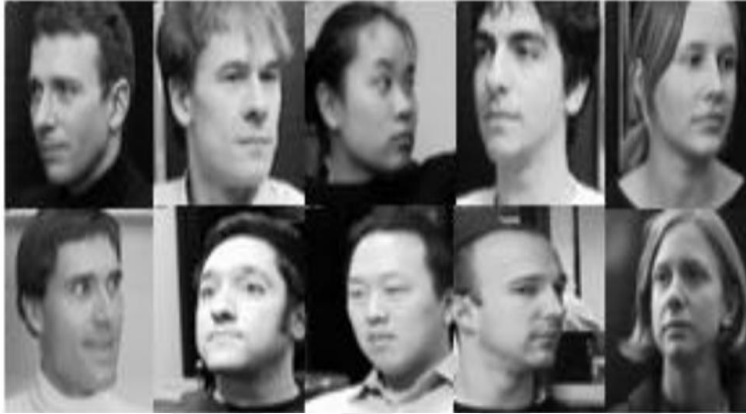
Figure 3.4. MIT CBCL face dataset

The correct classification accuracies on CBCL dataset are given in Table 3.3. Actually, all methods produced a reasonable accuracy value for CBCL method because the number of classes is just two. In other words, the dataset is not as challenging as the other datasets that we used in our previous experiments. As seen in Table 3.3, the highest accuracy was produced by ELM-LRF method. The accuracy is 98.34%. The second best accuracy was obtained with Pose invariant method where the accuracy is 95.40%. Finally, the worse result was produced by Original C2 features [14] method and the accuracy was 87.05%.

Table 3.3. Results for the MIT-CBCL dataset

Method	Testing Accuracy (%)
ELM-LRF	98.34
Pose invariant	95.40
Original C2 features [14]	87.05
MPCALDA [15]	88.53

4. CONCLUSIONS

This thesis investigates the local receptive fields of extreme learning machine (ELM-LRF) which logically outcomes from the chance of combinatorial hidden nodes and hidden neurons recommended in extreme learning machine concepts. According to extreme learning machine models, different kinds of local receptive fields can be tested and there exists local relationship among local receptive fields and random hidden neurons can be selected to set its weight. The local relations amongst input and hidden nodes might let the network to reflect local structures and combinatorial nodes additional present translational invariance addicted to the system. While the input weight is randomly produced and then orthogonalized in a directive to inspire a complete set of topographies. The output weights are intended logically in ELM-LRF as long as a simple deterministic answer. In the execution, we construct a detailed system for simplicity using the phase function to sample the resident influences and the square/square-root pooling construction to express the combinatorial nodes. Astonishingly random convolutional nodes can be measured an effective implementation of local receptive fields of extreme learning machine, while numerous of different types of the LRF is supported by extreme learning machine models. The experiment result of this thesis, first we tried three different face datasets to see the aim of ELM-LRF in face recognition and pass the iteration section while there is a lot of problems still stayed unsolved, the extreme learning machine and local Receptive fields got more advantage over those face datasets which applied in ELM-LRF. Now if we take an example from Table 3.3, there is a comparison between (LBPHS, LDPHS, POEMHS, FS-LBP, and ELM-LRF) algorithms, the experimental result showed that POEMHS algorithm can have best testing accuracy, which is 67.11%, and ELM-LRF was 66.11%, and from list of Table 3.3 and 3.4, the ELM-LRF has received very good testing accuracy among those algorithms applied to the same face datasets.

5. REFERENCE

[1] **Nissenbaum, H. (2009)**. Privacy in context: Technology, policy, and the integrity of social life. Stanford University Press.

[2] **W. W. Bledsoe**. The model method in facial recognition. Technical report pri 15, Panoramic Research, Inc., Palo Alto, California, 2002.

[3] **V. Blanz and T. Vetter**. Face recognition based on fitting a 3d morphable model. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(9):1063–1074, September 2003.

[4] **Brown, J. B., MacLeod, S. C., Macnaughtan, C., Smith, M. A., & Smyth, B.** (1968). A rapid method for estimating oestrogens in urine using a semi-automatic extractor. Journal of Endocrinology, 42(1), 5-15.

[5] **Vanzella, E., Cristiani, S., Fontana, A., Nonino, M., Arnouts, S., Giallongo, E., ... & Zaggia,** S. (2004). Photometric redshifts with the Multilayer Perceptron Neural Network: Application to the HDF-S and SDSS. Astronomy & Astrophysics, 423(2), 761-776.

[6] **A. Golstein, L. Harmon,** and A. Lest. Identification of human faces. Proceedings of the IEEE, 59:748–760, 1971.

[7] **T. Kenade**. Picture Processing System by Computer Complex and Recognition of Human Faces. PhD thesis, Kyoto University, November 1973.

[8] **J. Bigün, G. Chollet** Proceedings of the International Conferences on Audio- and Video-Based Person Authentication, 1997,1999.

[9] **Franka Gustin,Ivor Rendulic, Nikola Miskovic, Zoran Vukic** Proceedings of the International Conferences on Automatic Face and Gesture Recognition, 1995-1998.

[10] **Wechsler, H., Kakkad, V., Huang, J., Gutta, S., & Chen, V.** (1997, March). Automatic video-based person authentication using the RBF network. In International Conference on Audio-and Video-Based Biometric Person Authentication (pp. 85-92). Springer, Berlin, Heidelberg.

[11] **Yang, A. Y., Zhou, Z., Balasubramanian, A. G., Sastry, S. S., & Ma, Y**. (2013). Fast ℓ_{1}-Minimization Algorithms for Robust Face Recognition. *IEEE Transactions on Image Processing*, *22*(8), 3234-3246.

[12] **Messer, K., Kittler, J., Sadeghi, M., Marcel, S., Marcel, C., Bengio, S., ... & Srisuk, S**. (2003, June). Face verification competition on the XM2VTS database. In International Conference on Audio-and Video-Based Biometric Person Authentication (pp. 964-974). Springer, Berlin, Heidelberg.

[13] **P.J. Phillips, H. Moon, S. Rizvi, and P. Rauss**, "The FERET Testing Protocol," in Face Recognition: From Theory to Applications (H. Wechsler, P.J. Phillips, V. Bruce, F.F. Soulie and T.S. Huang, eds.), Berlin: Springer-Verlag, pp. 244-261, :1998.

[14] **K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre**, "XM2VTSDB: The Extended M2VTS Database," in Proceedings, International Conference on Audio-and Video-Based Person Authentication, pp. 72-77, 1999.

[15] **Chellappa, R., Wilson, C. L., & Sirohey, S.** (1995). Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, *83*(5), 705-741.

[16] **J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma,** "Robust face recognition via sparse representation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 2, pp. 210–227, Feb. 2009.

[17] **Tan.X B.Triggs**, ―Enhanced local texture feature sets for face recognition under difficult lighting conditions,‖ IEEETrans. Image Process.,vol.19, no.6, pp.1635–1650, Jun.2010.

[18] **L. Zhao and Y.H. Yang,** "Theoretical analysis of illumination in pcabased vision systems," Pattern Recognition, vol. 32, pp. 547-564, 1999.

[19] **Samaria, F. S., & Harter**, A. C. (2011, December). Parameterisation of a stochastic model for human face identification. In Applications of Computer Vision, 2011. Proceedings of the Second IEEE Workshop on (pp. 138-142). IEEE.

[20] **Hesher C., Srivastava A., and Erlebacher. A G.**, Novel technique for face recognition using range images. Seventh Int'l Symp. on Signal Processing and Its Applications, 2003.

[21] **P.K. Kalocsai, W. Zhao, and E. Elagin**, "Face Similarity Space as Perceived by Humans and Artifical Systems," in Proceedings, International Conference on Au-tomatic Face and Gesture Recognition, pp. 177-180, 1998.

[22] **Y. Le Cun.** Generalisation and network design strategies. Technical Report CRG-TR-89-4, Department of Computer Science, University of Toronto, 1989.

[23] **Y. Le Cun and Yoshua Bengio.** Convolutional networks for images, speech, and time series. In Michael A. Arbib, editor, The Handbook of Brain Theory and Neural Networks, pages 255–258. MIT Press, Cambridge, Massachusetts, 1995.

[24] **Y. Le Cun, B. Boser, J.S. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel**. Handwritten digit recognition with a backpropagation neural network. In D. Touretzky, editor, Advances in Neural Information Processing Systems 2, pages 396–404. Morgan Kaufmann, San Mateo, CA, 1990

[25] **Bai, L. L. C. Kasun,** Local Receptive Fields Based Extreme Learning Machine, IEEE Computational Intelligence Magazine, Vol. 10, Iss. 2, pp. 18-29, May 2015

[26] **Chao-Kuei Hsieh, Shang-Hong Lai, and Yung-Chang Chen**, ―An Optical Flow-Based Approach to Robust Face Recognition under Expression Variations‖, IEEE transactions on image processing, vol. 19, no. 1, jan 2010.

[27] **M.Kabir, T.Jabidand O.Chae**, ―A local directional pattern variance(LDPv) based face descriptor for human facial expression recognition,‖ in Proc.seventh IEEEInt.Conf.Adv. Video Signal Based Surv., Sep.2010, pp.526–532.

[28] **Raichle, M. E., & Mintun, M. A**. (2006). Brain work and brain imaging. *Annu. Rev. Neurosci.*, *29*, 449-476.

[29] **Gershenson, C.** (2003). Artificial neural networks for beginners. arXiv preprint cs/0308031.

[30] **Schmidhuber, J.** (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85-117.

[31] **Ciregan, D., Meier, U., & Schmidhuber, J.** (2012, June). Multi-column deep neural networks for image classification. In Computer vision and pattern recognition (CVPR), 2012 IEEE conference on (pp. 3642-3649). IEEE.

[32] **Matsuoka, K., & Yi, J**. (1991, November). Backpropagation based on the logarithmic error function and elimination of local minima. In Neural Networks, 1991. 1991 IEEE International Joint Conference on (pp. 1117-1122). IEEE.

[33] **Kim, Y.** (2014). Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.

[34] **F.Khellah**, ―Texture Classification Using Dominant Neighborhood Structure‖ IEEE transactions on image processing, vol. 20, no. 11, nov2011.

[35] **Huang, G. B., Zhu, Q. Y., & Siew, C. K.** (2006). Extreme learning machine: theory and applications. Neurocomputing, 70(1-3), 489-501.

[36] **C. P. Chen and C.-Y. Zhang,** "Data-intensive applications, challenges, techniques and technologies: A survey on big data," Inform. Sci., vol. 275, pp. 314–347, Aug. 2014.

[37] **Z.-H. Zhou, N. Chawla, Y. Jin, and G. Williams**, "Big data opportunities and challenges: Discussions from data analytics perspectives," IEEE Comput. Intell. Mag., vol. 9, pp. 62–74, Nov. 2014.

[38] **Y. Zhai, Y.-S. Ong, and I. Tsang**, "The emerging 'big dimensionality,'" IEEE Comput. Intell. Mag., vol. 9, pp. 14–26, Aug. 2014.

[39] **Husmeier, D., & Taylor, J. G.** (1998). Neural networks for predicting conditional probability densities: Improved training scheme combining EM and RVFL. *Neural Networks*, *11*(1), 89-116.

[40] **C. Garcia and M. Delakis**. Convolutional face finder: A neural architecture for fast and robust face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(11):1408–1423, 2004

[41] **Huang, G. B., Bai, Z., Kasun, L. L. C., & Vong, C. M**. (2015). Local receptive fields based extreme learning machine. IEEE Computational Intelligence Magazine, 10(2), 18-29.

[39] **Zhang, Z.** (1999). Feature-based facial expression recognition: Sensitivity analysis and experiments with a multilayer perceptron. International Journal of Pattern Recognition and Artificial Intelligence, 13(06), 893-911.

[40] **Rojas, R.** (1996). The backpropagation algorithm. In Neural networks (pp. 149-182). Springer Berlin Heidelberg.

[41] **Huang, G. B., Bai, Z., Kasun, L. L. C., & Vong, C. M**. (2015). Local receptive fields based extreme learning machine. IEEE Computational Intelligence Magazine, 10(2), 18-29.

[42] **Zheng, W., Zou, C., & Zhao, L.** (2004, August). Real-time face recognition using Gram-Schmidt orthogonalization for LDA. In Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on (Vol. 2, pp. 403-406). IEEE.

[43] **Lawrence, S., Giles, C. L., Tsoi, A. C., & Back, A. D. (1997).** Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks*, *8*(1), 98-113.

CIRRICULUM VITAE

ARASS MASOOD ISMAIL

Profile Summary

I am able to work on my own initiative or as part of a team. I am a self-motivated, flexible, enthusiastic and hardworking individual with knowledge and skills in IT and Business Management. I thrive on meeting new challenges. I wish to show my skills in teaching and research.

Key Skills

➢ Programming in various languages such as C++, c#, HTML and ASP.NET

➢ The neural network technology been used on my research for developing systems that is able to learn things like humans brain

➢ Networking Skills and familiar with Routing and Switching

➢ Multilingual : Kurdish, Arabic and English

➢ Great Organization Skills

Education

➢ 2012–2013 B.Sc, Soran University, First Class Honor's, Soran.Specialized in Computer Science

➢ 2007–2008 Higher Secondary, Bardarash Secondary School, First Class, Erbil. Specialized in Computer Science

Experience Vocational

➢ 2014-2015

Polytechnic University /College of Informatics

Working as assist programmer in IT Department.

➢ 2012–-2013 Assistant programmers, Soran University, Soran.

Taking lab lectures; Also doing some other tasks for example Working in university Resources, Giving university details to the student, example scientific debate, Help Lecturers Especially while they give exam to the students.

➢ 2011–2012 Network Engineer, Korektelecom company, Erbil.Managing IConnect system; Helping employers and solving them internet problem; Taking carry of system.

➢ 2010–2011 Teacher, Soran School, Soran.

Teaching given curriculum to year 2010 and year 2011, Organizing students and course material; Supervising and Setting Exams; Marking student exam papers.

➢ 2008–2009 Travel Agent, Travel Company, Erbil.

Booking ticket for clients; Organizing files and documents; Customer Services; Filling forms and sending them to Baghdad to the French Embassy.

Languages

Kurdish Mother tongue

English Conversationally fluent

Interests

- Reading - Writing

- Chess - Swimming

- Football