Chapter 2

Literature Review:

Pattern Recognition and its techniques
Abstract:

Associative memories can be implemented using networks with or without feedback. The goal of learning is to associate known input vectors with given output vectors. Contrary to continuous mappings, the neighborhood of a known input vector $x$ should also be mapped to the image $y$ of $x$. Noisy input vectors can then be associated with the correct output. The function of an associative memory is to recognize previously learned input vectors, even in the case where some noise has been added. The advantage of associative memories compared to this approach is that only the local information stream must be considered. This chapter includes review of literature on the related issues of my research work. I have gone through with the various literature resources to get the familiarity with the various image-processing techniques, pattern recognition algorithms, pattern recalling methods, feature extraction methods and evolutionary algorithms.
2.1 Introduction:

Since our early childhood, we have been observing patterns in the objects around us (e.g., toys, flowers, pets, and faces). Learning patterns also reinforces, and is reinforced by, the acquisition of language. A human can identify any object or we can say at the age of five years, most children can recognize digits and letters. Lower case characters, upper case characters, handwritten alphabets, machine printed or rotated images / characters are easily recognized by the young children [47]. Any image or figure of characters may be written on a cluttered background, on crumpled paper or may even be partially occluded. Pattern recognition is the task of recognizing patterns so that machines can observe the environment, learn to distinguish patterns of interest from their background, and make an appropriate, sound and reasonable decisions about the categories of the patterns. Pattern recognition is the research area that studies the operation and design of systems that recognize patterns in data. Application areas such as image analysis, character recognition, Speech recognition, man and machine diagnostics, person identification and industrial inspection are of tremendous practical significance and encompassed by pattern recognition. It has been well understood that these problem can be performed well and effortlessly by human brain. However, their solution using computer has, in many cases proved to be immensely difficult. In order to have the best opportunity of developing effective solutions, it is important to consider the various existing approaches and methods for pattern recognition with the machine. In this chapter we are discussing the general introduction of pattern recognition and various techniques or methods those are used for this purpose starting from conventional statistical inference to modern approaches of soft computing.
The best pattern recognizers in most instances are humans, yet we do not understand how humans recognize patterns. Ross [48] emphasizes the work of Nobel Laureate Herbert Simon whose central finding was that pattern recognition is critical in most human decision making tasks:

The concept of recognition is simple in the real world environment, but in the world of computer science, recognizing any object is an amazing feat. The functionality of the human brain is amazing; it is not comparable with any machines or software. Recognition is a basic property of all human beings; when a person sees an object, he or she first gathers all information about the object and compares its properties and behaviors with the existing knowledge stored in the mind. If we find a proper match, we recognize it [49]. Pattern recognition, as a subject, spans a number of scientific disciplines, uniting them in search for a solution to the common problem of recognizing the pattern of a given class and assigning the name of identified class. Pattern recognition is the categorization of input data into identifiable classes through the extraction of significant attributes of the data from irrelevant background detail. A pattern class is a category determined by some common attributes. Therefore, a pattern is the description of a category member representing a pattern class.

A pattern class is a family of patterns that shares some common properties. Watanabe [50] defines a pattern as opposite of a chaos; it is an entity, vaguely defined, that could be given a name. Pattern could be any static image, fingerprint image, a handwritten cursive word, a human face, or a speech signal. For example, some pattern recognition system applications include the following. In weather
prediction, input data are in the form of weather maps, and the output is a forecast. The symptoms serve as input data in medical diagnosis while disease identity serves as the output. The predicted market ups and downs are the desired output for stock market prediction when input data are the financial news and charts. Pattern recognition by machine involves techniques for assigning patterns to their classes automatically with as little human interventions is possible. Pattern recognition aims to classify data (patterns) based on either a priori knowledge or on statistical information extracted from the patterns. The patterns to be classified are usually groups of measurements or observations, defining points in an appropriate multidimensional space [51].

2.2 Pattern Recognition:

Pattern recognition aims to classify data (patterns) based on either a priori knowledge or on statistical information extracted from the patterns. The patterns to be classified are usually groups of measurements or observations, defining points in an appropriate multidimensional space [52]. A complete pattern recognition system consists of a sensor that gathers the observations to be classified or described; a feature extraction mechanism that computes numeric or symbolic information from the observations; and a classification or description scheme that does the actual job of classifying or describing observations, relying on the extracted features [53].

Automatic (machine) recognition, description, classification, and grouping of patterns are important problems in a variety of engineering and scientific disciplines such as biology, psychology, medicine, marketing, computer vision, artificial intelligence, and remote sensing. But what is a pattern? Watanabe [54] defines a
pattern as opposite of a chaos; it is an entity, vaguely defined, that could be given a
name. For example, a pattern could be a fingerprint image, a handwritten cursive
word, a human face, or a speech signal. Given a pattern, its recognition/classification
may consist of one of the following two tasks [58]:

Note that the recognition problem here is being posed as a classification or
categorization task, where the classes are either defined by the system designer (in
supervised classification) or are learned based on the similarity of patterns (in
unsupervised classification).

Pattern recognition problems may be logically divided into two broad
categories. First involves the study of pattern recognition capabilities of human
beings and other living organism, while the second involves the development of
theory and techniques for the design of device capable of performing a given
recognition task for a specific applications. The first area deals with the subjects of
psychology, physiology and biology, but the second area deals with the computer,
information science and artificial intelligence or machine learning methods. The act
of recognition with machine can be divided into two broad categories:

1. Concrete item recognition, it involves the recognition of spatial samples such
as fingerprints, weather maps, pictures and physical objects and the
recognition of temporal samples such as, waveforms and signatures.
2. Abstract item recognition, it involves the recognition of a solution to a
problem, an old conversation or argument.
A complete pattern recognition system consists of a sensor that gathers the observations to be classified or described; a feature extraction mechanism that computes numeric or symbolic information from the observations; and a classification or description scheme that does the actual job of classifying or describing observations, relying on the extracted features. The pattern recognition is concerned primarily with description and analysis or classification of measurements taken from physical or mental process [55].

The area of pattern recognition deals with the following:

- Character Recognition: Optical or hand written,
- Image Recognition,
- Voice Data Recognition,
- Speech Analysis,
- Man and Machine Diagnostics,
- Person Identification and Industrial Inspection.

During the last few years the researchers have proposed many mathematical approaches to solve the pattern recognition problems. Recognition strategies heavily depend on the nature of the data to be recognized. In the cursive case, the problem is made complex by the fact that the writing is fundamentally ambiguous as the letters in the word are generally linked together, poorly written and may even be missing. On the contrary, hand printed word recognition is more related to printed word recognition, the individual letters composing the word being usually much easier to isolate and to identify. As a consequence of this, methods working on a letter basis
Literature Review: Pattern Recognition and its techniques

(i.e., based on character segmentation and recognition) are well suited to hand printed word recognition while cursive scripts require more specific and/or sophisticated techniques. Inherent ambiguity must then be compensated by the use of contextual information. The available methods of pattern recognition may be categorized into four basic principles:

i. Statistical Methods; consisting the sub disciplines like discriminate analysis, feature extraction, error estimation, cluster analysis

ii. Structural Methods consisting grammatical inference and parsing

iii. Neural Networks based Methods

2.3 Statistical Methods of Pattern Recognition:

This is an approach to machine intelligence which is based on statistical modeling of data. In a statistical model, one applies probability theory and decision theory to get an algorithm. During the past there has been progress in theory and applications of 'Statistical Pattern Recognition' [56-60]. The three major issues encountered in the design of a statistical pattern recognition system are sensing, feature extraction, and classification. The primary issue is the representation of the input data which can be measured from the objects to be recognized and it is called sensing problem. Each measured quantity describes the characteristics of the pattern or objects.

The number of features of the pattern samples is usually very large. The features of the pattern samples are reduced by considering their salient characteristics. This process is referred to as feature extraction. Several
approaches for feature extraction have been proposed by the various researchers such as feature extraction by moment’s invariants [61], feature extraction by autoregressive models [62], and feature extraction by KL transformation [63].

The last issue of the statistical pattern recognition is the pattern classification or development of the classifier. The pattern classifier is defined as a device or a process that sorts the given data into identifiable categories and classes. The pattern classification is an information transformation process, i.e., the classifiers transforms relatively large set of mysterious data into a smaller set of useful data [60]. Trainable classifier is one that can improve its performance in response to the information it receives as a function of time. Training is a process by which the parameters of the classifiers are adjusted. The classifier is trained using the reduced pattern samples. It is often assumed that the pattern samples of a given class occupy a finite region in a pattern space and it is called a class region.

A good classification is the main object of recognition system. Generally, the statistically pattern recognition problems fall into two main categories: supervised classification (discriminate analysis) problems and unsupervised classification (clustering) problems. When the samples have known classification (labeled samples) then the classification is called the supervised otherwise it is unsupervised [64, 65]. So in the supervised classification, the data is labeled; and the problem is to design a classifier to predict the class label for any given sample. In the unsupervised classification, data are not labeled and to classify the data into groups with features that distinguish one group from another.
These basic approaches can be organized into the tree structure as shown in Figure 2.1.

![Figure 2.1: Various Approaches in Supervised Classification](image)

Ahmad et al [66] made a comparison between statistical and neural pattern recognition techniques and realize that how neural techniques are much far better than statistical techniques. They made a comparison between Discriminant Analysis (DA) and Principal Component Analysis (PCA) which were used for pattern recognition, and are statistical techniques. Discriminant Analysis has few problems of small sample size and huge data dimensions. To overcome these problems, pattern recognition was also done using Generalized Regression Neural Network (GRNN) and Back-propagation Neural Network (BPNN) techniques. Neural networks proved better results than statistical methods.
2.4 Neural Network Based Methods:

The pattern recognition approaches discussed so far are based on direct computation through machines. Direct computations are based on statistical analysis techniques. The neural approach applies biological concepts to machines for pattern recognition. The outcome of this effort is invention of artificial neural networks. Neural networks can be viewed as massively parallel computing systems consisting of an extremely large number of simple processors with many interconnections. Neural network models attempt to use some organizational principles (such as learning, generalization, adaptively, fault tolerance, distributed representation, and computation) in a network of weighted directed graphs in which the nodes are artificial neurons and directed edges (with weights) are connections between neuron outputs and neuron inputs. The main characteristics of neural networks are that they have the ability to learn complex nonlinear input-output relationships, use sequential training procedures, and adapt themselves to the data. The most commonly used family of neural networks for pattern classification tasks [67] is the feed-forward network, which includes multilayer perceptron and Radial-Basis Function (RBF) networks. These networks are organized into layers and have unidirectional connections between the layers. Another popular network is the Self-Organizing Map (SOM), or Kohonen-Network [68], which is mainly used for data clustering and feature mapping. The learning process involves updating network architecture and connection weights so that a network can efficiently perform a specific classification/clustering task. The increasing popularity of neural network models to solve pattern recognition problems has been primarily due to their seemingly low dependence on domain-specific knowledge (relative to model-based and rule-based approaches) and due to the availability of efficient learning algorithms. Neural
networks provide a new suite of nonlinear algorithms for feature extraction (using hidden layers) and classification (e.g., multilayer perceptrons). In addition, existing feature extraction and classification algorithms can also be mapped on neural network architectures for efficient (hardware) implementation.

Due to the explosive growth of digital technologies, ever increasing visual data are created and stored. Nowadays, visual data are as common as textual data. There is an urgent need of effective and efficient tool to find visual information on demand. A large amount of research has been carried out on image retrieval (IR) in the last two decades [73]. In general, IR research efforts can be divided into three types of approaches. The first approach is the traditional text based annotation. In this approach, images are annotated manually by humans and images are then retrieved in the same way as text documents [69-72]. However, it is impractical to annotate a huge amount of images manually. Furthermore, human annotations are usually too subjective and ambiguous.

The second type of approach focuses on content based image retrieval (CBIR), where images are automatically indexed and retrieved with low level content features like color, shape and texture [73-74, 76-82]. However, recent research has shown that there is a significant gap between the low level content features and semantic concepts used by humans to interpret images. In addition, it is impractical for general users to use a CBIR system because users are required to provide query images. The third approach of image retrieval is the automatic image annotation (AIA) so that images can be retrieved in the same way as text documents [75-98, 99,100]. The main idea of AIA techniques is to automatically learn semantic
concept models from large number of image samples, and use the concept models to label new images. Once images are annotated with semantic labels, images can be retrieved by keywords, which is similar to text document retrieval. The key characteristic of AIA is that it offers keyword searching based on image content and it employs the advantages of both the text based annotation and CBIR. There are several surveys on broad CBIR research in literature [101-106], and a survey on broad semantic IR techniques is given by Liu et al. [107].

In image classification and retrieval, images are represented using low level features. Because an image is an unstructured array of pixels, the first step in semantic understanding is to extract efficient and effective visual features from these pixels. Appropriate feature representation significantly improves the performance of the semantic learning techniques. While both global and region based image representations are used in the existing image retrieval techniques, the trend is towards using region based features. Region based feature extraction needs prior image segmentation while global features are directly calculated from the whole image.

Image segmentation is usually the first step to extract region based image representation. The segmentation algorithm divides images into different components based on feature homogeneity. A number of segmentation approaches exist in the literature, such as grid based, clustering based, contour based, model based, graph based, and region growing based method. This section provides a brief review of segmentation methods commonly used in AIA. For a comprehensive segmentation review, readers are referred to [108].
Because automatic image segmentation is a difficult task, many techniques simplify this task using grid based approach to roughly segment images into blocks [76, 78, 81, 83-85, 87, 109-110]. Visual features are then extracted from these blocks. Block based approach takes little computations; however, this simple technique does not describe the semantic components in images well. A single block often consists of parts of visually different objects. Furthermore, it is difficult to determine the size of blocks for image representation. Therefore, region features are usually not accurate. If appropriately applied, it can be used in domain specific applications, e.g., medical image classification [111].

Clustering algorithms, like k-means, are used to cluster pixels into different groups [112,77,113,114], with each group identifying a region. In most cases, an image is first partitioned into blocks of size 44 pixels. Color and/or texture features are extracted for each block. Then, k-means is applied to cluster the block feature vectors. A region is formed with the pixels belonging to blocks of the same cluster. The major issue with this approach is that it needs to predefine the number of segments based on heuristics. An inappropriate choice of k may yield poor results. The other issue is that the algorithm assumes data are in spherical clusters so that the mean values are near the cluster centres. This assumption, however, is usually not true.

The main idea of contour based segmentation is to evolve a curve around an object. The evolution stops when the curve coincides with the boundary of an object. Unlike the cluster based segmentation algorithm, contour based segmentation algorithms do not need the prior assumption of the number of clusters [115-117].
The underlying problem in this approach is the dependency on accurate edge detection which is subject to noise. Therefore, it often needs human to define rough boundary outline which makes the approach to be applicable only to specific domain, e.g., image processing tools.

Segmentation algorithms based on statistical models have also been proposed [118-120]. Among them, Blobworld [118] is widely used [118, 112, 122]. In Blobworld, each pixel is represented by an 8 dimensional feature vector of color, texture and position. An image pixel is then modeled as a random variable with Gaussian mixture distributions. The number of regions and Gaussian parameters are calculated using expectation maximization (EM) algorithm. Once the model parameters are found, the pixel–region relationship is calculated using the posterior probabilities. The pixel–region relationship is used to determine the image segmentation. One of the major issues with this approach is that the computation is very expensive because the EM is an optimization algorithm.

Shi and Malik [112] propose a graph based segmentation algorithm known as normalized cut (NCut). The NCut method represents an image as a graph where vertices are image pixels and the edge weights represent the feature similarities between pixels. Image segmentation then becomes a graph partitioning problem. The idea is to partition the vertices of the graph into disjoint sets so that the total similarity between different sets is minimized. Each set is regarded as region. As the number of pixels in an image is large, there are exponential numbers of possible partitions of the graph. As a result, it is computationally expensive to find the optimal partition. Tao et al. [116] improves the NCut by pre-segmenting images

-(63)-
using mean shift algorithm [119]. Instead of using pixels, the regions of the initial segmentation are used as vertices in the NCut algorithm. Hence, the computational cost is reduced, and the performance is more robust. The basic NCut is based on color features only. Malik et al. [123] extend it to incorporate texture features.

The widely used JSEG [124] algorithm is a region growing approach. It groups pixels or smaller regions into larger regions. At first, pixel colors of the image are quantized into a number of classes and pixels in the image are replaced with the color class labels. A class map is formed and region growing is followed on the class map. Pixels with more homogeneous neighbors are assumed to be interior pixels of possible regions. These pixels are selected as candidate seed points and regions are grown around these seed areas. As this method looks for color and texture homogeneity, the segmented regions have highly homogeneous characteristics. It has been widely used in image retrieval [96, 74, 123, 125].

Image segmentation is a complex issue and a large research topic. Segmentation performance usually depends on applications. For image retrieval purpose, the region boundary does not have to be accurate as long as the region is homogenous. However, regions from segmentation are usually contaminated with segments from neighboring regions. This problem can be overcome by a clean-up post-processing [74].

Image segmentation plays an important role in the screening of medical imaging. In the last decades, fuzzy segmentation methods, especially the fuzzy c-means algorithm, have been widely used in the image segmentation and such a
success chiefly attributes to the introduction of fuzziness for the belongingness of each image pixel [126]. This allows for the ability to make the clustering methods able to retain more information from the original image than the crisp or hard segmentation methods [127].

2.5 Pattern Recognition using GA and Neural network system:

Since first attempts to combine GA and NN started in the late 1980s, other researchers have joined the movement and created a flood of journal articles, technical reports etc. A broad variety of problems have been investigated by different GANN approaches, such as face recognition, animals [128], classification of the normality of the thyroid gland [129], color recipe prediction [130] and many more. Also, a variety of different encoding strategies have been implemented [131]. A method has been developed for Off-line Handwritten character recognition with Genetic Algorithm [132].

Lots of work has been already done on the evolution of neural network with hybrid GA [132, 133]. The majority of implementations of the GAs are derivatives of Holland’s [134] innovative specification. Evolution has been introduced in NNs at three levels: connection weights, architectures and learning rules. The evolution rules have not yet been subjected to a similar study, but the literature on the subject is mounting. The evolution of a network’s connection weights is an area of curiosity and successfully implemented in [135]. In this approach the GA is used in the feed forward neural network for evolving the population of the weights is evaluated from the fitness evaluation function. The least mean square error function is used for the evaluation of individual weight population. The fittest weights are used for further
computation and participate in the classification. This process will continue until the required classification is achieved. The final selected weights represent the optimized strength of the connection in the network architecture, which is suitable for the classification of all presented patterns. This strength represents the convergence of the weights for the desired classification. There is also a possibility that more than one population of weights is generating the correct classification for every presented training pattern. [136, 137].

In [138], an improved GA is used for training a three-layer FNN with switches at its links. Both the nonlinear mapping and the network architecture can be learned. The weights of the links govern the input-output mapping, while the switches of the links govern the network architecture. In the genetic backpropagation (G-Prop) method [139], the GA selects the initial weights and changes the number of neurons in the hidden layer through the application of five specific genetic operators, namely, mutation, multipoint crossover, addition, elimination and substitution of hidden units. The BP, on the other hand, is used to train from these weights. This makes a clean division between global and local search. This strategy attempts to avoid Lamarckism. Behaviors between parents and their offspring are linked by various mutations, such as partial training and node splitting. The evolved neural network is parsimonious by preferring the node/connection deletion operations to the node/connection addition operations. The hybrid training operator that consists of a modified BP with adaptive learning rates and the SA is used for modifying the connection weights. After the evolution, the best evolved neural network is further trained using the modified BP on the combined training and validation set.
There are also studies on evolving node activation functions or learning rules, since different activation functions have different properties and different learning rules have different performance [140]. For example, the learning rate and the momentum factor of the BP algorithm can be evolved [141], and learning rules evolved to generate new learning rules [142]. EAs are also used to select proper input variables for neural networks from a raw data space of a large dimension, that is, to evolve input features. The activation functions can be evolved by selecting among some popular nonlinear functions such as the Heaviside, sigmoid, and Gaussian functions. A neural network with evolutionary neurons has been proposed in [143].

A new genetic-based prototype system for localizing 2-D compound objects inside plane images was introduced and tested in the localization of LP symbols. The results were encouraging and a new approach for solving the LP detection problem relying only on the geometrical layout of the LP symbols was experimentally proved [144]. Conventional feature selection techniques usually demand many samples to estimate statistics accurately. In addition, they are usually based on an exhaustive process for finding the best set of features, and in this case, they are time demanding, and their CPU processing time exponentially increases as the number of bands (features) increases [145]. To this extent, a new generation of feature selection techniques is based on evolutionary optimization methods, since they are not based on an exhaustive process and can lead to a conclusion in a faster way. In addition, by considering an efficient fitness function for these methods, they can handle high-dimensional data with even a limited number of training samples (ill-posed
Literature Review: Pattern Recognition and its techniques

situations). In particular, the genetic algorithm (GA) and particle swarm optimization have gained significant attention from researchers.

2.6 Conclusion:

The objective of this chapter was to introduce the general introduction of pattern recognition and its various techniques. Pattern recognition problems and its various existing solutions using neural network techniques and other well known techniques are discussed. Pattern recognition is a fast-moving and proliferating discipline. It is not easy to form a well-balanced and well-informed summary view of the newest developments in this field. It is still harder to have a vision of its future progress.

As discussed, the neural networks have the ability to learn complex nonlinear input-output relationships, use sequential training procedures, and adapt themselves to the data. Neural networks have also several advantages such as, unified approaches for feature extraction & classification and flexible procedures for finding good, moderately nonlinear solutions. The neural networks are now well established as an important technique for solving pattern recognition problems, and indeed there are already many commercial applications of feed-forward neural networks in routine use. The complexity and characteristics of handwritten pattern recognition is also discussed in neural networks based method framework with the problem of its implementation.

The various modifications and improvements are also considered and reported. Various techniques of feature extractions those have been investigated in
literature are reported with their merits and demerits. In the last the hybrid evolutionary techniques are reviewed. This review started from the discussion of Genetic algorithm with its capabilities for searching. There are various works base on GA with neural network techniques have reported for the task of pattern recognition. This review has also considered the various methods those have been reported in the literature for the combination of GA with neural network on various parameters. The most common method to apply the GA for the evolution of population of weights has reported. It has investigated from the review that the hybrid evolutionary algorithm is more effective to explore the solution by escaping the local minimum even for the large search space. This literature review highlights the successive development in pattern recognition with various approaches of artificial neural networks. It has also indicated that the recognition of hand written curve script is an interesting and very popular domain area for the pattern recognition task. This review process completed with the effectiveness of hybrid evolutionary algorithm for the handwritten curve script recognition tasks.