Volume 3, Issue 4, April 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

# Classification in Pattern Recognition: A Review

**Priyanka Sharma**
*Department of Computer Science and Engineering*
*PEC University of Technology, Chandigarh, India*

**Manavjeet Kaur**
*Department of Computer Science and Engineering*
*PEC University of Technology, Chandigarh, India*

***Abstract: Pattern Recognition is one of the very important and actively searched trait or branch of artificial intelligence. It is the science which tries to make machines as intelligent as human to recognize patterns and classify them into desired categories in a simple and reliable way. This review paper introduces the basic concepts of pattern recognition, the underlying system architecture and provides the understanding of various research models and related algorithms for classification and clustering.***

***Keywords: Pattern recognition, supervised learning, unsupervised learning, PCA, KPCA, ANN, LDA***.

## I. INTRODUCTION

The process of recognizing patterns and classifying data accordingly has been gaining interest from a long time and human beings have developed highly sophisticated skills for sensing from their environment and take actions according to what they observe. So a human can recognize the faces without worrying about the varying illuminations, facial rotation, facial expressions, and facial biometrical changes and also occluded face images. But if the point of implementing such recognition artificially came, then it becomes a very complex task. The fields of artificial intelligence have made this complex task possible by making machines as intelligent as human to recognize patterns in varying environmental conditions. Such a branch of artificial intelligence is known as pattern recognition. Pattern Recognition provides the solution to a lot of problems that fall under the category of either recognition or classification, such as speech recognition, face recognition, classification of handwritten characters, medical diagnosis etc.

### *A. Pattern*

A pattern is a set of objects or phenomena or concepts where the elements of the set are similar to one another in certain ways or aspects. There are various definitions proposed for the term *pattern.*

"A pattern is essentially an arrangement. It is characterized by the order of the elements of which it is made, rather than by the intrinsic nature of these elements," is a definition given by Norbert Wiener [16].

Watanabe [14] defines a pattern as "opposite of a chaos; it is an entity, vaguely defined, that could be given a name".

"It can also be defined by the common denominator among the multiple instances of an entity. For e.g., commonality in all fingerprint images defines the fingerprint pattern; thus, a pattern could be a fingerprint image, a handwritten cursive word, a human face, a speech signal, a bar code, or a web page on the Internet"[20]. The following are some of the examples of patterns.

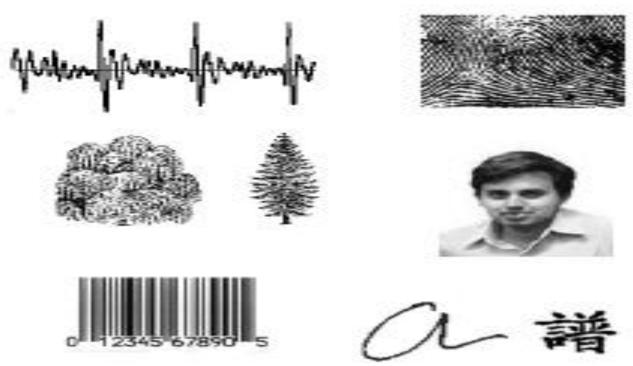

**Fig. 1 Example of Patterns: Finger Print, Sound Wave, Tree, Face, Bar Code & Character Image [1]**

### *B. Pattern category*

It is a collection of similar, not necessarily identical objects. Often, individual patterns may be grouped into a category based on their common properties; the resultant is also a pattern and is often called a pattern category.

 

*C. Pattern recognition*

It is defined as the study of how machines can observe the environment, learn to distinguish various patterns of interest from its background, and make reasonable decisions about the categories of the patterns. During recognition, the given objects are assigned to a prescribed category.

*D. Pattern recognition system*

The design model of a pattern recognition system essentially involves the following three steps [1][4]:

1) *Data acquisition and preprocessing*: Here the data from the surrounding environment is taken as the input and given to the pattern recognition system. The raw data is then preprocessed by either removing noise from the data or extracting pattern of interest from the background so as to make the input readable by the system.
2) *Feature extraction*: Then the relevant features from the processed data are extracted. These relevant features collectively form entity of object to be recognized or classified.
3) *Decision making*: Here the desired operation of classification or recognition is done upon the descriptor of extracted features.

Block diagram of a pattern recognition system is shown in figure 2.

## II. Pattern Recognition Models

There are four basic models followed in pattern recognition; these are statistical model, syntactical or structural model, template matching model and neural network based model.

*A. Statistical model*

It is the most intensively used model in pattern recognition systems because it is the simplest to handle. The statistical pattern recognition systems are based on statistics and probabilities. Here each pattern is described in terms of feature sets. Feature sets are chosen in such a way that different patterns occupy non-overlapping feature space. The effectiveness of the feature set is determined by how well patterns from different classes can be separated i.e., there is a proper interclass distance. After performing the analysis of the probability distribution of a pattern belonging to a certain class, a decision boundary is determined [23]. Here the patterns are projected to some pre-processing operations to make them suitable for training purposes. Features are selected after analyzing the training patterns. System learns from the training patterns and adapts itself to recognize or classify the unknown test patterns. Feature measurement is done while testing, i.e., distance between the patterns is determined in the statistical space and then these feature values are presented to learnt system and in this way classification is performed.

Block diagram of statistical model for pattern recognition is shown in figure 3.

*B. Syntactic model*

These models are also named as structural models for pattern recognition and are based on the relation between features. Here the patterns are represented by structures which can take into account more complex relations between features unlike the numerical feature sets used in statistical pattern recognition models. Also the patterns used in this model forms a hierarchical structure composed of sub-patterns. In this model, the patterns to be recognized are called primitives and the complex patterns are represented by the inter-relationship formed between these primitives and the grammatical rules associated with this relationship [3]. In syntactic pattern recognition, a similarity is associated between the structure of patterns and the syntax of a language. The patterns are the sentences belonging to a language, primitives are the alphabet of the language, and using these primitives, the sentences are generated according to the grammar. Thus, the very complex patterns can be described by a small number of primitives and grammatical rules [2][4]. This approach is considered to be an appealing model in pattern recognition because, in addition to classification, it also provides a description of how from the primitives the given pattern is constructed due to its hierarchical structure. This paradigm has been used in situations where the patterns have a definite structure which can be captured in terms of a set of rules [2]. The implementation of a syntactic model approach, however, leads to many difficulties because of the segmentation of noisy patterns (to detect the primitives) and the inference of the grammar from training data. This may yield a combinatorial explosion of possibilities to be investigated, demanding a very large training sets and huge amount of computational efforts [19].

*C. Template matching model*

This is a widely used model in image processing to determine the similarity between two samples, pixels or curves to localize and identify shapes in an image. In this model, a template or a prototype of the pattern to be recognized is available. Each pixel of the template is matched against the stored input image while taking into account all possible position in the input image, each possible rotation and scale changes. In visual pattern recognition, one compares the template to the input image by maximizing the spatial cross-correlation or by minimizing a distance: that provides the matching rate. After calculating the matching rate for every possibility, select the largest one which exceeds a predefined threshold. It is a very expensive operation while dealing with big templates and using large sets of images. Also it does not work efficiently in the presence of distorted patterns [1][2]. The given figure explains about the steps performed in template matching process. Here the first figure is the input image and a small portion of it acts as a test template. Then matching is performed and the position of template is marked.

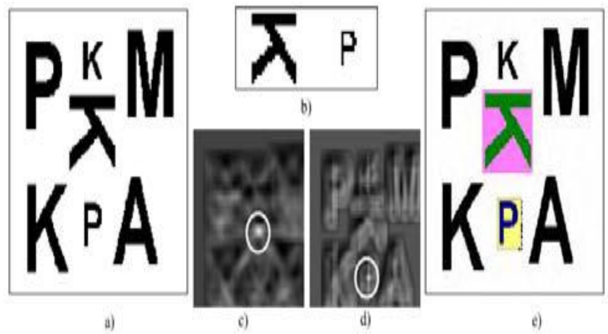


**Fig 4.Example of template matching model for pattern recognition [1]**

*D. Neural network model*

Neural networks can be viewed as a parallel computing systems consisting of an extremely large number of simple processors with many interconnections between them. Typically, a neural network or to be more specific, an artificial neural network (ANN) is a self-adaptive trainable process that is able to learn and resolve complex problems based on available knowledge. An ANN-based system behaves in the same manner as how the biological brain works; it is composed of interconnected processing elements that simulate neurons. Using this interconnection, each neuron can pass information to another. Artificial Neural network models attempt to use some organizational principles such as learning, generalization, adaptivity, fault tolerance and distributed representation, and computation in the network of weighted directed graphs in which the artificial neurons forms the nodes of the model and the directed edges (with weights) are connections between neuron outputs and neuron inputs[2][4]. The weights applied to the connections results from the learning process and indicate the importance of the contribution of the preceding neuron in the information being passed to the following neuron [1]. The main characteristics of all the neural networks are that they possess the ability to learn complex nonlinear input-output relationships, use sequential training procedures, and adapt themselves to the data. The following diagram is a two layer neural network with one input layer constituting of three neurons and one output layer with two neurons and corresponding weights are assigned in between them.

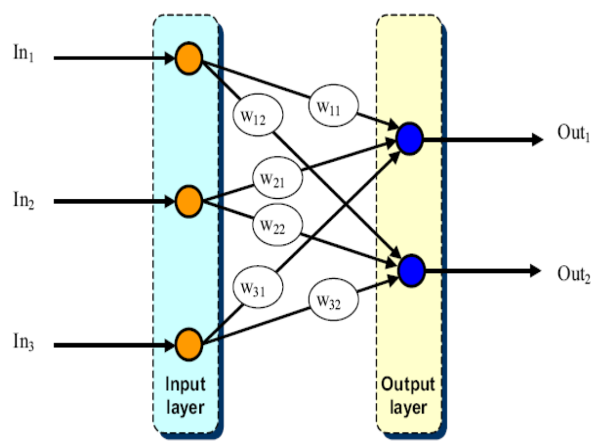


**Fig 5.Example of an artificial neural network [7]**

Table 1 highlights the important characteristics of the above explained pattern recognition models.

| MODEL | REPRESENTATION | RECOGNITION FUNCTION | TYPICAL CRITERION |
|---|---|---|---|
| Statistical | Features | Discriminant Function | Classification Error |
| Syntactic or Structural | Primitives | Rules, Grammar | Acceptance Error |
| Template Matching | Sample, pixel, Curves | Correlation, distance measures | Classification Error |
| Neural Networks | Samples, pixels, Features | Network Function | Mean square Error |

## III. PATTERN RECOGNITION ALGORITHMS

The field of pattern recognition has been explored widely by a number of researchers who as a result have developed various algorithms. The design pattern of all these algorithms consists of three basic elements, i.e., data perception, feature extraction and classification. There are various different techniques to implement these three basic elements. So which technique is chosen for each element in design cycle defines the algorithm characteristic of the pattern recognition algorithm.

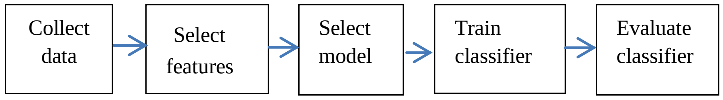


Fig. 6 Basic pattern recognition model

This is the design cycle of a basic pattern recognition model. Algorithms for pattern recognition depend on the type of label output, on whether learning is supervised or unsupervised [17].

### *A. Supervised learning*

Supervised learning is a process of allotting a function to some desired category as learnt from supervised training data. Here the training data consist of a set of training examples where each set consist of a pair consisting of an input object and a desired output value. A supervised learning algorithm learns from this training pair relationship and produces an inferred function.

In simple terms, in supervised learning, there is a teacher who provides a category label or cost for each pattern in the training set which is used as a classifier.

So basically a supervised learning method is used for ***classification*** purpose. In the given figure, the input image consist of a mixture of two alphabets, i.e., A and B. Then the classification algorithm classifies the input to two different categories.

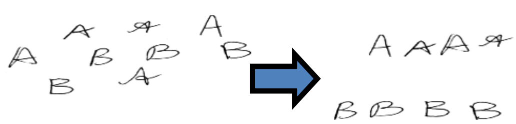

**Fig. 7 Example of supervised learning**

Here a set of combined input is classified using supervised learning approach.

### **B.** *Unsupervised learning*

Un-supervised learning can be defined as the problem of trying to find out the hidden structure in an unlabeled data set**.** Since the examples given to the learner are unlabeled, each algorithm itself classifies the test set [1]. In simple terms, here no labeled training sets are provided and the system applies a specified clustering or grouping to the unlabeled datasets based on some similarity criteria. So an unsupervised learning method is used for ***clustering.*** Here the input consists of some unlabeled values whose distinguishing feature is initially not known. The following input consists of such a combination with all values technically same but still its clusters are formed using some metric which is different for each algorithm.

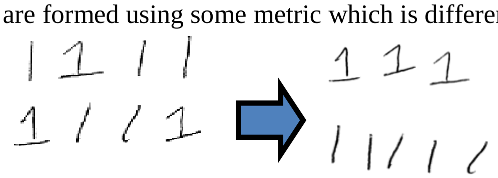

Fig 8. Example of unsupervised learning

Here clusters are formed in the output.

## IV. CLASSIFICATION ALGORITHMS

### *A. Linear discriminant analysis (LDA)*

It is used to find a linear combination of features which characterizes or separates two or more classes of objects or events. LDA is a parametric approach in supervised learning technique. It was initially used for dimensionality reduction and feature extraction, and later moved for classification purpose also.LDA easily handles the cases where the within-class frequencies are unequal and their performances had been examined on randomly generated test data. Thus it maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal separability. Linear discriminant analysis has a close relation with Principal Component Analysis (PCA). Both methods are used for dimensionality reduction. LDA have been proven better algorithm when compared with PCA.The prime difference between LDA and PCA is that PCA does more of feature classification and LDA does data classification. In working with PCA, the location of the original data set changes when transformed to a totally different space whereas LDA doesn't change the location but only tries to provide more class separability and draw a decision region between the given classes.

*B. Quadratic Discriminant Analysis (QDA)*

It is used in machine learning and statistical classification to separate measurements of two or more classes of objects or events by a quadric surface. It is a more general version of a linear classifier. QDA is a parametric approach in supervised learning which models the likelihood of each class as a Gaussian distribution, then uses the posterior distributions to estimate the class for a given test point. The Gaussian parameters for every class can be estimated from training points with maximum likelihood (ML) estimation. The simple Gaussian model assumption is best suited to cases when one does not have much information to characterize a class, i.e. , if there are too few training samples to infer much about the class distributions. Also, when the number of training samplesis small compared to the number of dimensions of each training sample, the ML covariance estimation can be ill-posed. There exists some solutions to resolve this ill-posed estimation; one is to regularize the covariance estimation and another is to use Bayesian estimation [1].

*C. Maximum entropy classifier (multinomial logistic regression)*

In statistics, a maximum entropy classifier model is a regression model which generalizes logistic regression by allowing more than two discrete outcomes. This forms a model that is used to predict the probabilities of the different possible outcomes of a categorically distributed dependent variable, given a set of independent variables (which may be real-valued, categorical-valued etc.).The actual goal of the multinomial logistic regression model is to predict the categorical data. Maximum entropy classifiers are commonly used as an alternative to Naive Bayes classifiers because they do not require statistical independence of the independent variables (commonly known as features) that serve as the predictors. This algorithm may not be appropriate to learn large number of classes since it is slower than for a Naive Bayes classifier. Multinomial logistic regression is a particular solution to the classification problem that assumes that a linear combination of the observed features and some problem-specific parameters can be used to determine the probability of each particular outcome and the best values of the such parameters for a given problem are usually determined from some training data.

*D. Decision trees*

It is considered to be a decision support tool that uses a tree-like structure or model of decisions and all its possible consequences. It is one way to display an algorithm. These trees are basically used in operations research, mostly in decision analysis, to help identify a strategy most likely to reach a goal.In this process, a decision tree and the closely related influence diagram is used as a visual and analytical decision support tool where the expected values of competing alternatives are calculated.Decision trees are a simple, but very powerful form of multiple variable analysis [24]. The trees provide unique capabilities which act to be supplement, complement, and substitute for

- Traditional statistical forms of analysis (such as multiple linear regressions)
- A lot of data mining tools and techniques (such as neural networks)
- The recently developed multidimensional forms of reporting and analysis found in the field of business intelligence

The decision trees are produced by algorithms which identify various ways of splitting the data set into branch-like segments. These segments form an inverted decision tree which starts with a root node at the top of the tree. Each node starting from root contains the name of field which is also called object of analysis. The decision rule is discovered based on a method that extracts the relationship between the object of analysis (that serves as the target field in the data) and one or more fields that serve as input fields to create the branches or segments. The values of the input field are used to estimate the likely value of the target field which can also be termed as an outcome, response, or dependent field or variable. Once the relationship is found, then one or more decision rules can be derived which describe the relationships between inputs and targets. Then these decision rules can be used to predict the values of new or unseen observations which contain values for the inputs, but might not contain values for the targets.

*E. Kernel Estimation & K-nearest neighbor*

In the field of pattern recognition, the k-nearest neighbor algorithm (k-NN) is a method for classifying objects based on closest training examples in the feature space. K-NN is a type of example-based learning, or lazy learning where the function is only approximated locally and all the computation is deferred until classification. This algorithm is one of the simplest machine learning algorithm in which an object is classified using a majority vote of its neighbors and the object is then assigned to the class which is most common amongst its k-nearest neighbors. Here the neighbors are taken from a set of objects for which the correct classification is known. These neighbors can be assumed as a training set for this algorithm, though no explicit training step is required. The learning in this model is based on storing all the training instances which corresponds to points in an n-dimensional Euclidean space along with their class labels and classification is delayed till a new instance is arrived. As the new unlabeled query instance or vector arrives, the classification is performed by assigning the label which is most frequent among the k-training samples nearest to that query point.

There are some variations that can be performed on this algorithm. These variations start with 1-NN where k=1 and the object is simply assigned to the class of its nearest neighbor. Then we have k-NN approach where the value of k is chosen randomly. Here we find k closest training points to the test instance according to some metric (mostly used metric is the Euclidean distance) and then perform classification operation. The best choice of k generally depends on the data itself.

 

However larger value of k reduces the effect of noise on classification but makes the boundaries between classes less distinct. So a good choice of k is required which can be achieved by some heuristic technique called cross-validation.

*F. Naive Bayes classifier*

Naive Bayes classifier is a simple, probabilistic and statistical classifier which is based on Bayes theorem (from Bayesian statistics) with strong (naive) independence assumptions and maximum posteriori hypothesis. As Bayesian classifiers are statistical in nature, they can predict the probability of a given sample belonging to a particular class. The underlying probability model to this classifier can be termed more appropriately as an "independent feature model" because a naive Bayes classifier assumes that the effect of an attribute value on a given class is independent of the values of the other attributes. Such an assumption is called class conditional independence. It is made to simplify the computation involved and, in this sense, is considered "naive". We can explain this classifier with a small example. A fruit is considered to be an apple if it is red in color, round in shape, and around 5" in diameter. Although these features depend on each other or upon the existence of the other features, a naive Bayes classifier takes all of these properties to independently contribute to the probability that this fruit is an apple. The naive Bayes classifier is trained using a supervised learning approach that just requires consideration of each attribute in each class separately. So the training in naive Bayes classifier is considered to be very easy and fast. To estimate the parameters in naive Bayes model it uses the principle of maximum likelihood method in many practical applications. Testing in this algorithm is also very straightforward and simple;just look the tables and calculate conditional probabilities with normal distributions. The advantage of Naive Bayes model is that it only requires a small amount of training data to estimate the parameters, i.e., means and variances of the variables which are necessary for classification.

*G. Artificial Neural Networks*

It is an interconnected network of a group of artificial neurons.An artificial neuron can be considered as a computational model which is inspired by the natural neurons present in human brain. Unlike natural neurons, the complexity is highly abstracted when modeling artificial neurons. These neurons basically consist of *inputs* (like synapses), which are further multiplied by a parameter known as *weights* (strength of each signals), and then computed by a mathematical function which determines the *activation* of the neuron. After this there is another function that computes the *output* of the artificial neuron (sometimes in dependence of a certain *threshold*). Thus the artificial networks are formed by combining these artificial neurons to process information.

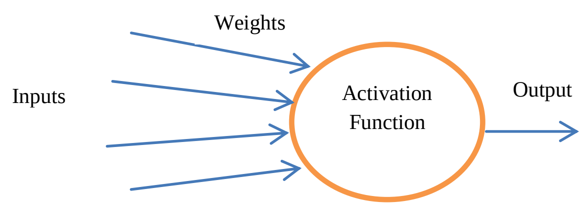


Fig 9.An artificial neuron

We can train ANN for best matched solution; ANN can perform fuzzy matching and provides the optimal solution. It also acts as a classifier in pattern recognition. It falls under the category of supervised learning where the model initially learns from the training data set and then classifies the test image using the learnt knowledge.

*H. Support Vector Machine*

A Support Vector Machine (SVM) performs classification by constructing an *N*-dimensional hyperplane that optimally separates the data into two categories. A support vector machine (SVM) is used in computer science for a set of related supervised learning methods that analyze input data and learns from it and then use it for performing classification and regression analysis. The standard SVM is a two-class SVM which takes a set of input data and predicts the possible class, for each input, among the two possible classes the input is a member of, which makes it a non-probabilistic binary linear classifier. Given the set of training examples where each one of them is marked as belonging to one of the two categories, the SVM training algorithm builds a model that assigns new examples into one category or the other. SVM is an efficient method of finding an optimal hyperplane for separating non-linear data also. Presently, the traditional two-class SVM is also used in multiclass classification where the data to be classified may belong to any one class among a number of classes.

## IV. CLUSTERING ALGORITHMS

*A. Hierarchical Clustering*

It is a process used in data mining concept where it can be defined as a method of cluster analysis which works to build a hierarchy of clusters. It is a widely used data analysis tool. The idea behind hierarchical clustering is to build a binary tree of the data that successively merges similar groups of points and visualizing this tree provides a useful summary of the data.
Hierarchical clustering strategies generally fall into two types:

 

1) *Agglomerative*: This is a "bottom up" approach of hierarchical clustering where each observation starts with one single cluster, and then pairs of clusters are merged as one move up the hierarchy. In Agglomerative clustering, each level of the resulting tree is a segmentation of the data. Hence the algorithm results in a sequence of grouping and then it is up to the user to choose a natural clustering from this sequence.
2) *Divisive*: This is a "top down" approach of hierarchical clustering where all the observations start in one cluster, and splits are performed recursively as one moves down the hierarchy. Here a dissimilarity measure is required between the sets of observations to decide which all clusters should be combined for agglomerative clustering, or where a cluster should split for divisive clustering. Mostly in hierarchical clustering, this measure is achieved by use of an appropriate metric and a linking criterion which specifies the dissimilarity of sets as a function of the pairwise distances of observations in the sets.

### *B. K-means Clustering*

As a process employed in data mining, k-means clustering is defined as a method of cluster analysis which aims to partition n different observations into k different clusters in which each observation belongs to the cluster with the nearest mean. Although this problem is computationally very difficult and has been put under the NP hard problem set, there are efficient heuristic algorithms that are commonly employed and converge quickly to a local optimum. Such algorithms are similar to the expectation-maximization algorithm for mixtures of Gaussian distributions via an iterative refinement approach employed by both algorithms.

### *C. KPCA (Kernel Principle Component Analysis)*

It is an extension of principal component analysis (PCA) or may also be termed as a nonlinear form of PCA. Using this form of PCA one can efficiently compute principal components in a very high dimensional feature spaces related to input space by some non-linear mapping using techniques of kernel methods and functions. Particularly for clustering, KPCA can be used to construct a hyperplane that divides the 'n' points into arbitrary clusters by making them almost always linearly separable in d >=n dimensions. Also this nonlinear kernel PCA can be used for simple pattern recognition with a linear classifier with much better recognition rates in comparison to simple PCA. Along with this, the computational complexity of KPCA does not grow with the dimensionality of the feature space it is working on.

## VI. DISCUSSION AND CONCLUSION

It has always been difficult to decide which algorithm is best to classify patterns with least computational effort, least time and maximum and best results. In this review paper, various categories of pattern recognition algorithms are discussed. Pattern recognition field has a wide range of applications in the field of classification, clustering, regression, sequence labeling and parsing among which this paper reviews the algorithms of the most applied field on pattern recognition, i.e., classification and clustering. The classification approach to pattern recognition uses labeled training set with which it classifies the test unlabeled data to the desired category. In contrast to this, the clustering algorithms don't have a labeled set. They use some other metric like Euclidean distance to put the test set into correct cluster.

## References

[1] M. Parasher, S. Sharma, A.K Sharma, and J.P Gupta, "Anatomy On Pattern Recognition,*" Indian Journal of Computer Science and Engineering (IJCSE*), vol. 2, no. 3, Jun-Jul 2011.

[2] Anil K. Jain, Robert P.W. Duin, and Jianchang Mao, "Statistical pattern recognition-A review," *IEEE transactions on Pattern Analysis and Machine Intelligence,* vol. 22, no. 1, January 2000.

[3] Jie Liu, Jigui Sun, Shengsheng Wang, "Pattern Recognition: An overview", *IJCSNS International Journal of Computer Science and Network Security*, vol. 6, no. 6, June 2006.

[4] SeemaAsht and RajeshwarDass, "Pattern Recognition Techniques: A Review", *International Journal of Computer Science and Telecommunications,* vol. 3, issue 8, August 2012.

[5] Vinita Dutt, VikasChadhury, Imran Khan, "Different Approaches in Pattern Recognition",*Computer Science and Engineering. 2011*; 1(2): 32-35.

[6] Majida Ali Abed , Ahmad Nasser Ismail and ZubadiMatizHazi, "Pattern recognition Using Genetic Algorithm",*International Journal of Computer and Electrical Engineering*, Vol. 2, No. 3, June, 2010.

[7] Ahmad, T.,Jameel, A., Ahmad, B., "Pattern recognition using statistical and neural techniques", *International Conference onComputer Networks and Information Technology (ICCNIT),* 2011.[8] Mohammad S. Alam,Mohammad A. Karim, "Advances in Pattern Recognition Algorithms, Architectures and Devices," *Optical Engineering*, Vol. 43 No. 8, August 2004.

[9] Sebastien ´ Gadat,Laurent Younes, "A Stochastic Algorithm for Feature Selection in Pattern Recognition", *Journal of Machine Learning Research* 8 (2007) 509-547.

[10] Mu Zhu, Wenhong Chen, John P. Hirdes, Paul Stolee, "The K-nearest neighbor algorithm predicted rehabilitation potential better than current Clinical Assessment Protocol", *Journal of Clinical Epidemiology* 60 (2007) 1015-1021.

 

[11] SantoshSrivastava, Maya R. Gupta andBela A. Frigyik, “Bayesian Quadratic Discriminant Analysis”, *Journal of Machine Learning Research* 8 (2007) 1277-1305.

[12] M. Saberi, A. Azadeh, A. Nourmohammadzadeh and P. Pazhoheshfar, “Comparing performance and robustness of SVM and ANN for fault diagnosis in a centrifugal pump”, *19th International Congress on Modelling and Simulation, Perth, Australia*, 12–16 December 2011.

[13] Durgesh K. Srivastava, LekhaBhambhu, “Data Classification using Support Vector Machine”,*Journal of Theoretical and Applied Information Technology*© 2005 - 2009 JATIT.

[14] S. Watanabe, “Pattern Recognition: Human and Mechanical”, *New York: Wiley*, 1985.

[15] Duda, Richard O., Hart, Peter E., and Stork, David G., “ Unsupervised Learning and Clustering”, *Chapter 10 in Pattern classification (2nd edition), p. 571, New York, NY: Wiley, ISBN 0-471-05669-3*

[16] R. C. Gonzalez, “Object Recognition”, *in Digital image processing, 3rd ed. Pearson*, August 2008, pp. 861-909

[17] Introduction to pattern recognition; Web source by Wikipedia.

[18] Sholom M. Weiss and IoannisKapouleas, “An Empirical Comparison of Pattern Recognition,Neural Nets, and Machine Learning Classification Methods”.

[19] Perlovsky, L.I. (1998), ‘Conundrum of combinatorial complexity’. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 20, no. 6.

[20] Anil k Jain, Robert P.W Duin, “Introduction to pattern recognition”, *The Oxford Companion to the Mind, Second Edition*, Oxford University Press, Oxford, UK, 2004, 698-703.

[21] S. P. Shinde and V.P.Deshmukh, “Implementation of Pattern Recognition Techniques and Overview of its Applications in various areas of Artificial Intelligence”, *International Journal of Advances in Engineering & Technology*, Vol. 1, Issue 4, pp. 127-137,Sept 2011.

[22] Vapnik, V., “ The Nature of Statistical Learning Theory”, *Springer*, 2010.

[23] NavdeepKaur and Usvir Kaur, “Survey of Pattern Recognition Methods”, *International Journal of Advanced Research in Computer Science and Software Engineering* ,Volume 3, Issue 2, February 2013.

[24] Barry de Ville, “Decision Trees for Business Intelligence and Data Mining”, *SAS Enterprise Miner*, Jan 2007.


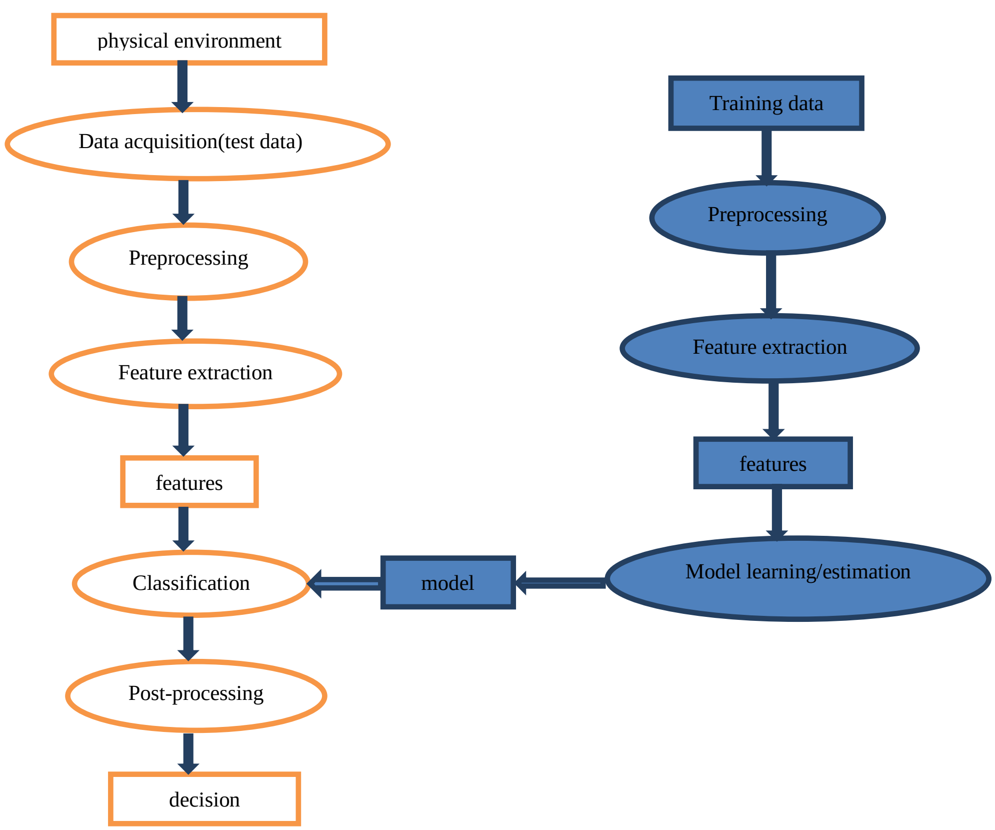


Fig 2.Block diagram of a pattern recognition system [1]

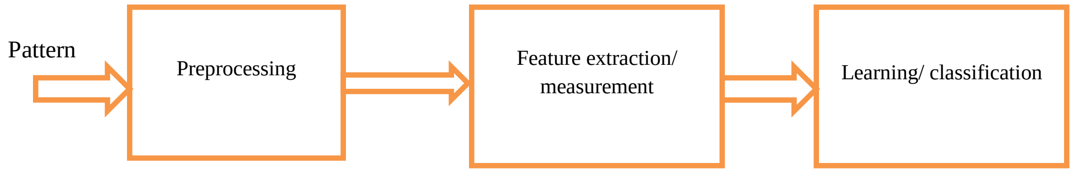


Fig 3.Block diagram of a statistical model of Pattern recognition

 