CHAPTER 6

MACHINE LEARNING APPROACHES FOR HUMAN RECOGNITION FROM PROPOSED RUSH FEATURES

6.1 INTRODUCTION

As a human being, one needs to take decisions and to make judgments in some situations. Classification is a human activity where a decision or a forecast has to be made according to the currently available information present in the system under consideration. Pattern recognition refers to the construction of a classification procedure for assigning a new case to a class out of a set of pre-defined classes according to certain attributes or features. It is also called discrimination or supervised learning. Classification plays a prominent role in any object recognition system where it is carried out by any of the machine learning approaches.

Machine Learning (ML) refers to the science of automatic computing to learn tasks from a series of instances. The rationale behind this chapter is to discuss the concepts of machine learning and to formulate a proper machine learning approach for recognizing human based on the proposed RUSH features extracted from ear biometric images. This chapter presents the machine learning approaches proposed for human recognition based on ear biometrics using the proposed RUSH features.
6.2 MACHINE LEARNING

Learning is the process of converting experience gained over a certain period into commendable expertise or knowledge. Several learning procedures are followed in all contexts of learning. With respect to the field of computer vision, several learning algorithms are followed to gain knowledge in order for a particular task to be completed. The input to a learning algorithm is training data which represents experience and the output is some knowledge required to perform the task for which the application is built. This learning process can be made automated which will then be called ‘Machine Learning’.

6.2.1 Characteristics of machine learning

Machine Learning (ML) plays a very important role in any automated system which requires certain knowledge to be extracted out of some experience given as input in the form of training data. Arthur Samuel is considered to be a pioneer of Machine Learning and was an expert in computer gaming and artificial intelligence, who coined the term ‘Machine Learning’ in 1959 during his tenure at IBM.

Machine learning algorithms should possess the following characteristics.

- Learning by memorization

This is the foremost characteristic of machine learning where in the automated system memorizes the results of training data and produces matching results if the system is tested with a data from the training it has already undergone. This phenomenon is said to be ‘Learning by memorization’.
• **Learning by generalization**

   Although learning by memorization characteristic is useful, it lacks the ability to label or categorize an unseen data – the data which is not a part of training, the system underwent. As in human, a successful learner should be able to apply generalization in case of unseen data. This phenomenon is said to be ‘Learning by generalization’.

• **Adaptability**

   Machine learning programs have the ability to adapt to the inputs which are applied to the system in which ML programs run. This is essential because many complex tasks may be time variant and user dependent i.e. such tasks may change with respect to time and also change from one user to another.

**6.2.2 Need for machine learning**

There are so many reasons to deploy machine learning algorithms to solve complex tasks rather than solving these tasks using programmable computers with latest processors which are listed below.

• ML algorithms are used to solve tasks which are performed routinely by humans / animals. These algorithms utilize ML programs, programs that learn from their experience i.e. training. Few such examples are object recognition, image analysis and understanding etc.,

• ML algorithms are employed for solving complex tasks which are beyond the capabilities of human. These tasks include extraction of meaningful information from medical archives, astronomical data, and weather prediction. In all these cases,
abundant information is present but it is beyond human capabilities to process them. Learning is done to detect meaningful patterns in large datasets with the help of computers with latest processors.

6.2.3 Types of machine learning

ML techniques (Bishop CM 2006) can be classified into many categories depending upon the following four parameters.

- Supervised and Unsupervised Learning

Learning in ML algorithms involves an interaction between the learner and the environment and the ML algorithms are classified into two types based on the nature of interaction, namely supervised and unsupervised learning.

Supervised learning corresponds to the learning in which the training data or ‘experience’ contains significant information. Also, sometimes, the expertise acquired from training examples may be required to predict information in unseen test scenario. In the latter case, the environment which interacts with the learner, acts as a supervisor by providing extra information. Thus, the two ML characteristics such as learning by memorization and learning by generalization are utilized in supervised learning. The proposed ear biometric recognition methodology using RUSH features is successful in recognizing human based on ear biometrics by incorporating supervised learning for training the proposed biometric system with the extracted RUSH features.

Unsupervised learning corresponds to the learning in which there is no difference between training data and test data. The learner processes the
input data to produce a summarized or compressed version of the input data. Eg. Clustering data.

Reinforcement Learning is an intermediate of both supervised and unsupervised learning techniques where the training data contains more information than the test data and the learner has to predict even more information for test data.

- **Active and Passive Learning**

  ML paradigms can be classified into active and passive learning depending on the degree of interaction of the learner with environment. In active learning, the learner interacts with the environment during training by sending queries or by conducting experiments. In passive learning, the learner is merely an observer, observing the information provided by the learning environment (supervisor) and does not influence the environment.

- **Adversarial and Statistical Learning**

  Adversarial learning corresponds to the worst case scenario where the input of the learning system is generated by an adversarial teacher. If learning is successfully performed with an adversarial teacher (learning environment), then it can be very well performed with any odd learning environment. Statistical learning corresponds to the ML scenario where the learner’s experience gained from training data, is generated by a random process.

- **Online and Batch learning**

  This categorization of ML is based on whether the learner learns experience from the entire training dataset at once or in sequential order of steps. The former is said to be batch learning and the latter is said to be online
learning. Online learning is commonly used in various applications of ML where it is not computationally efficient to train over the entire dataset.

The advantages of online learning are as follows

1. Computationally faster and memory space efficient
2. Easier to implement
3. More general framework

6.3 PATTERN CLASSIFICATION WITH INSTANCE BASED CLASSIFIERS

Pattern classification plays a vital role in object and/or human recognition systems requiring the knowledge acquired from ever increasing data in this informative world to be recognized and utilized positively. Thus, pattern recognition (i.e. pattern classification) techniques come under the umbrella of machine learning terminology. Pattern recognition can very well be performed with a variety of ML techniques such as Support Vector Machines (SVM), Artificial Neural Networks, Deep Learning networks and so on. But instance based classifier is the simplest pattern classification technique which classify feature vectors based on some distance function. The extracted RUSH features are to be classified using the simplest distance based classifier in the proposed ear biometric model.

In order to employ machine language algorithms for a successful pattern classification task, ML algorithms should define an instance space first.

Let X denotes an instance space defined for the machine learning algorithm and it contains the set of all instances which in our investigation
corresponds to the RUSH features extracted from the ear biometric modalities acquired during both enrolment and recognition phases.

\[ X = \{ x_1, x_2, x_3, \ldots \ldots \ldots, x_d \} \]

Instance based classifiers are classifiers which classify unknown instances by relating the unknown instance to a known instance based on some distance function. The distance function can be Euclidean distance, City block distance, Manhattan distance and so on.

Some instance based classifiers include simple distance based classifiers based on these distance functions (nearest neighbor classifiers), KNN classifiers and so on. Instance based classifiers do not collect any information from the training dataset during enrolment or training phase. Learning by memorization, an important characteristic of machine learning does not take place unless such a memorization becomes unavoidable for classification. So, instance based classifiers are said to be Lazy learners unlike feed forward networks where learning by memorization takes place during enrolment phase and which are said to be Eager learners. These instance based classifiers carry out classification through learning by generalization.

6.3.1 Nearest Neighbour classifiers

For any instance to be classified, the following steps are followed.

1. Locate a nearest neighbor for the unknown instance to be classified in instance space.
2. Identify the class label of the located known neighbor of the unknown instance.
3. Label the unknown instance with the class label of the located neighbor.
This approach is called Nearest Neighbor classifier. The distance function used to identify the nearest neighbor can be Euclidean distance, Mahalanobis distance, City Block distance etc.

Let $X_i = [x_{i1}, x_{i2}, x_{i3}, \ldots, x_{id}]$ represents an $i^{th}$ instance or a feature vector from training dataset say, $S$ in our present investigation. Here, ‘d’ is the dimensionality of the instance. Let $Q=[q_1, q_2, q_3, \ldots, q_d]$ represents a query instance or feature vector to be classified. The dimensionality of $Q$ is same as that of all the instances in training dataset. Various distance functions used in nearest neighbor classifiers (Bishop CM 2006) are illustrated as follows.

- **Euclidean distance function**
  \[
  D_E(X_i, Q) = \sum_{j=1}^{d} (X_{ij} - Q)^2
  \]  
  \(6.1\)

- **Cityblock (or Manhattan) distance function**
  \[
  D_{CB}(X_i, Q) = \sum_{j=1}^{d} |(X_{ij} - Q)|
  \]  
  \(6.2\)

- **Mahalanobis (or Statistical) distance function**
  \[
  D_{M}(X_i, Q) = \sum_{j=1}^{d} [(X_{ij} - Q)^t S^{-1}(X_{ij} - Q)]
  \]  
  \(6.3\)

where $S$ corresponds to the Covariance matrix of training dataset and is calculated as

\[
S = \frac{1}{d} \sum_{j=1}^{d} [(X_{ij} - \mu)(X_{ij} - \mu)^t]
\]  
\(6.4\)

where $\mu$ is the mean of the training dataset.

The nearest neighbor classifiers using all the above distance functions are implemented in MATLAB (R2013a) version 8.1 for the
proposed ear recognition framework for classifying RUSH features of dimensionality 285 corresponding to sub image block of UMRT as 8 x 8 as in Table 4.1.

Though the nearest neighbor classifier is simple to use, it has certain drawbacks such as

- Lack of robustness
- High degree of local sensitivity
- High susceptibility to noise in the training dataset

6.3.2 K Nearest Neighbour classifiers

KNN classifier is a variant of the nearest neighbour classifier and is used to overcome the drawbacks of the simple nearest neighbor classifier.

For any instance to be classified, the following steps are followed,

1. Locate ‘K’ nearest neighbors, where K > 1 for the unknown instance to be classified in instance space.
2. Identify the class label of located ‘K’ neighbors of the unknown instance.
3. Apply majority voting to determine the outcome of class labeling i.e. the class label of the nearest neighbor to which the unknown instance is to be mapped.
4. Label the unknown instance with the outcome of majority voting.

There are two possible cases for ‘K’. If K = 1, then the classifier corresponds to the simple nearest neighbor classifier. If K = n where n
corresponds to the size of the training dataset, the KNN classifier reduces the worst case in which all unknown instances belong to the class most frequently represented in the training dataset.

The suitable value of the parameter ‘K’ which corresponds to the number of nearest neighbors in KNN classifier, is usually found out by trial and error. In our discussion also, the value of K is determined by trial and error and the possible values of K are furnished in Chapter 7.

6.4 PERFORMANCE CURVES FOR BIOMETRIC RECOGNITION

The biometric system can be built either as verification or identification system depending upon the application. Biometric data to be trained and tested is put into three datasets namely training, testing and evaluation datasets.

Training dataset consists of the biometric modalities of all candidates registered for the biometric application. All biometric modalities are acquired under uniform illumination conditions, subject to similar stages of pre-processing and the proposed RUSH features are extracted from these images.

Test dataset consists of biometric modalities of both registered as well as unknown subjects. The purpose of preparing this test dataset is to assess the performance of the investigated biometric recognition system and to evaluate performance measures. Higher are the values of desired performance measures, more practical is the biometric system.
Evaluation dataset consists of biometric modalities of registered as well as unknown subjects whose biometric data is not present in test dataset. In short, the evaluation and test datasets are said to be mutually exclusive.

The various performance charts (curves) and performance metrics (Fawcett T 2006) are shown in Figure 6.1.

![Figure 6.1 Performance curves and metrics for a biometric system](image)

The performance of a machine language algorithm especially, pattern recognition or pattern classification task is evaluated by plotting three important performance curves namely Receiver Operating Characteristic (ROC) curve, Cumulative Match Characteristic (CMC) curve and Expected Performance Characteristic (EPC) curve. The significance of these performance evaluation curves are described in the following subchapters.
6.4.1 Receiver Operating Characteristics (ROC) curve

Receiver Operating Characteristic (ROC) curve is a technique used to visualize, organize and evaluate the performance of a pattern classifier. ROC curves are conceptually simple and are of very much use among the machine learning community. These curves were initially used in the field of signal detection to portray the tradeoffs between hit rates and false alarm rates of classifiers. The utilization of ROC curves have been extended to the field of medical diagnostics. With the widespread deployment of biometric systems for human authentication, the performance curves such as ROC, CMC and EPC curves are being used to evaluate the performance of biometric systems. A classification model or a classifier can be regarded as the mapping of a set of instances in test data set to a set of predicted classes. Different thresholds are applied to the estimate of each instance’s class membership probability to predict the class membership. Consider a classifier model with a set of instances in the test dataset. For every instance, there are four possible outcomes:

a) If the instance is positive and it is classified correctly as positive, the outcome is labelled as ‘True positive’.

b) If the instance is negative and it is classified correctly as negative, the outcome is labelled as ‘True negative’.

c) If the instance is positive and it is classified incorrectly as negative, the outcome is labelled as ‘False negative’.

d) If the instance is negative and it is classified incorrectly as positive, the outcome is labelled as ‘False positive’.

For a classifier with a set of instances and four possible outcomes as depicted above, a matrix called confusion matrix is constructed. The
confusion matrix otherwise called (contingency table) consists of disposition of the set of instances in the test dataset. The numeric entries along the main diagonal of confusion matrix represent the correct decisions made in classifying the genuine users as genuine (True Positives) and imposters as imposters (True negatives). The numbers along off diagonal of the matrix represents the confusion exist between classes in the form of errors namely True Positives and True Negatives.

Figure 6.2 Confusion matrix of Receiver Operating Characteristic curve

In the confusion matrix shown in Figure 6.2, P corresponds to Total positives and N corresponds to Total negatives.

True positive rate of a classifier is given by TP rate = \[
\frac{\text{Positives Correctly classified (TP)}}{\text{Total Positives (P)}}
\]

False positive rate of a classifier is given by FP rate = \[
\frac{\text{Negatives Incorrectly classified (FP)}}{\text{Total Negatives (N)}}
\]

Specificity = \[
\frac{\text{True Negatives}}{\text{False Positives} + \text{True Negatives}} = 1 - \text{FP rate}
\]
Accuracy = \[ \frac{TP + TN}{P + N} \]

where TP corresponds to number of Positives correctly classified and TN refers to the number of Negatives incorrectly classified as Positives. P and N are the total number of positives and negatives respectively.

Accuracy of a diagnostic test or the recognition process is measured by the area under ROC curve. If the area under ROC curve is 1, then it represents a perfect test; an area of 0.5 represents a worthless test.

Table 6.1  Grading of Classification Accuracy based on Area under ROC curve

<table>
<thead>
<tr>
<th>Sl.No.</th>
<th>Area under ROC curve</th>
<th>Category of Classification Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9 – 1.0</td>
<td>Excellent (A)</td>
</tr>
<tr>
<td>2</td>
<td>0.8 – 0.9</td>
<td>Good (B)</td>
</tr>
<tr>
<td>3</td>
<td>0.7 – 0.8</td>
<td>Fair (C)</td>
</tr>
<tr>
<td>4</td>
<td>0.6 – 0.7</td>
<td>Poor (D)</td>
</tr>
<tr>
<td>5</td>
<td>0.5 – 0.6</td>
<td>Fail (F)</td>
</tr>
</tbody>
</table>

The accuracy of a diagnostic test is classified into five categories according to the traditional academic point system (Fawsett T et al. 2006) as in Table 6.1.

ROC curve is a 2D plot in which true positive rate is plotted along y axis against false positive rate along x axis. This curve gives the relative tradeoffs between true positives and false positives. Figure 6.3 depicts ROC curves of different discrete classifiers labeled A,B,C,D and E. A point, say ‘x’ in ROC space is said to be better than another point ‘y’ in the ROC space if the point x is to the northwest of the point y where TP rate is higher or FP rate is lower or both. In this aspect, the classifier ‘D’ stands to be the perfect classifier out of these 5 classifiers and the point (0,1) represents perfect
classification. The classifier ‘A’ is said to be more conservative than the classifier ‘B’. The classifiers ‘C’ and ‘E’ are poor classifiers.

Figure 6.3 ROC curves of 5 different discrete classifiers

In a biometric verification system, the degree of similarity between the biometric feature of each subject in test dataset is arrived with the biometric feature of all enrolled persons in the training dataset. The error rates are given as False Acceptance Rate (FAR) and False Reject Rate (FRR).

False Acceptance Rate (FAR) or False Match Rate (FMR) corresponds to the fraction of impostor scores which exceed a preset threshold say, T and the False Reject Rate (FRR) refers to the fraction of genuine scores which fall below the preset threshold, T; Genuine Acceptance Rate (GAR) corresponds to the fraction of genuine scores which exceed the preset threshold T and is calculated as GAR = 1 - FRR (Jain AK et al. 2007).

In the context of biometrics, ROC curves are used to summarise the performance of a biometric verification system. The ROC curve of a biometric verification system is obtained by plotting GAR vs FMR by varying
the values of threshold, $T$ after calculating genuine & imposter scores. In a typical biometric verification system, the ROC curve is plotted with Genuine Accept Rate against False Match Rate for various values of thresholds set while comparing match scores in the Decision module of a biometric recognition system (Struc V et al. 2009 & 2010). The ROC curve of a typical biometric verification system is shown in Figure 6.4.

![ROC Curve](image)

**Figure 6.4 ROC curve of a typical Biometric Verification system**

In some cases, the ROC curve is also obtained by plotting FAR Vs FRR for different values of threshold $T$. In our present investigation, ROC curves are drawn by plotting FAR Vs FRR for different values of threshold.

6.4.2 Cumulative Match Characteristic (CMC) curve

Another important performance curve for a biometric recognition system is the Cumulative Match Characteristic (CMC) curve. As ROC curve is associated with the performance of a verification system, the performance of an identification system is summarised with the help of CMC curves. Such a 1:N biometric identification system ranks all N candidates registered during
enrolment phase, with respect to a test (probe) biometric query. The ranking of candidates is done by sorting the matching scores based on the chosen distance measure (Struc V et al. 2010). CMC curves give a measure of how well the biometric identification system ranks the biometric modalities of enrolled persons with respect to ‘unknown’ probe biometric modality. Identification Accuracy refers to the probability that a test biometric image is identified correctly at least at rank-k. The CMC curve is plotted for Identification Accuracy (Rate) against Rank values. The

![CMC Curve](image)

**Figure 6.5 CMC curve of a typical Biometric Identification system**

CMC curve of a typical biometric identification system is portrayed in Figure 6.5. As the rank is plotted for Identification Accuracy (Rate) Vs Rank values, the starting point in this figure corresponds to Rank-1 Identification Rate or Rank-one Recognition Rate.

Rank-one Recognition rate plays a vital role to evaluate the performance of the biometric identification system. Higher the Rank-one Recognition Rate, better the performance of the biometric identification system is.
The CMC curves of two different biometric identification systems are portrayed in Figure 6.6 (a) and 6.6 (b). The CMC curve in Figure 6.6 (a) has a Rank-one Recognition Rate (Identification Accuracy) of 82% and that in Figure 6.6(b) has a Rank-one Recognition Rate of 55%. The CMC curve of Figure 6.6(a) is better than the CMC curve of Figure 6.6(b).

![CMC Curve](image)

(a)

![CMC Curve](image)

(b)

**Figure 6.6 CMC curves of two different identification systems**

In our present investigation, CMC curve is obtained by plotting Identification Accuracy (Rate) against various possible Rank values.

### 6.4.3 Expected Performance Characteristic (EPC) curve

The third important performance curve of the biometric system is the Expected Performance Characteristic (EPC) curve which is an a priori curve by plotting Half Total Error Rate (HTER) against a parameter $\alpha$. 
α ∈ [0,1] which balances between FAR and FRR;

HTER is calculated as the average of FAR and FRR on evaluation dataset and is threshold dependent.

6.5 SUMMARY

This chapter has elaborated the basic concepts and characteristics of machine learning while clearly elucidating the advantages and disadvantages of various ML techniques. The suitability of ML approaches such as KNN classifier and simple distance based classifiers for classifying the proposed RUSH features is discussed in this investigation. The significance of various performance measures such as Rank-one Recognition Rate and performance curves such as ROC, CMC and EPC curves in the context of biometric recognition systems is also explored. The procedure to generate these performance curves for the proposed ear biometric recognition system is also lucidly presented. The performance measures and curves for the proposed ear biometric recognition system are obtained in Chapter 7.