

**Name of Candidate: Ms. Gauri Katiyar**

**Name of Supervisor: Dr. Shabana Mehfuz**

**Department: Department of Electrical Engineering, JMI, New Delhi**

**Title: Hybrid Intelligent System for Handwritten Character Recognition**

#### **ABSTRACT**

Since the advent of the digital computer, machine simulation of human function has been a very challenging and fascinating research area in the field of image processing and pattern recognition. The motivation for the study in this area is that researchers are interested in developing an intelligent system which has human like skills with performance comparable to human beings. The advancement of computational paradigm has resulted in modern and machine oriented outlook which leads to betterment of society by using biological inspired methods. The interconnection between computer science and other information processing disciplines is taken care of by these technologies. Handwriting Recognition is the mechanism for converting the handwritten text into a notational representation. It is a special problem in the domain of pattern recognition and machine intelligence. Systems for recognizing machine printed text originated in the late 1950s and have been in widespread use on desktop computers since the early 1990s. In the early 1990s, image processing and pattern recognition are efficiently and effectively combined with artificial intelligence and statistical technique that is Hidden Markov Model (HMM). Off-line handwritten character recognition continues to be an active research area towards exploring the newer techniques because it has various applications such as postal sorting, bank cheque amount reading, and official document reading.

The problem of off-line handwritten character recognition has been investigated by many researchers over a long time. In spite of recent advances, recognition of handwritten character is still a challenging problem due to the large degree of variability of writing style, different size and orientation angle of the characters, and even now there is no single approach that solves it both efficiently and accurately.

Computer based pattern recognition is a process that involves several sub-processes, including pre-processing, feature extraction, feature selection, and classification. Feature extraction is the estimation of certain attributes of the target patterns. Selection of the right set of features is the most crucial and complex part of building a pattern recognition system. Feature selection is a process of selecting only significant features from a large database to create subset of the original database. This process retains the original feature without changing the feature's domain. The main objectives of the feature selection are: to reduce database dimensionally, remove irrelevant features, reduce time needed to learn the classification function, increase accuracy and to keep only important features that give comprehensive understandings for all variables. Demand of feature selection is currently rising because of the expanding size of databases in various applications.

To improve the performance of the off-line handwritten character recognition system either the performance of the classifier has to be improved or better feature extraction techniques and/or feature selection techniques need to be explored. We have presented a methodology for hybrid feature extraction and a Genetic Algorithm (GA) based approach for optimal selection of feature subset along with an adaptive Multi layer perception (MLP) as pattern classifier. Adaptive nature in the classifier is achieved by implementing a function for selection of best architecture of MLP during the feature selection and classification phases. We have extracted seven feature sets based on moment features, distance –based feature, geometrical feature and local features . The effectiveness of the method is tested on the problem of off-line handwritten character recognition to address the problem of diversity in style, size and shape, which can be found in handwriting produced by different writers. All the handwritten data considered

here are unconstrained alphabets to avoid the process of segmentation. Selection of features also plays an important role in improving the performance of the system. The novelty of the proposed recognition system is that by hybridization of the feature extraction techniques and randomly selecting the features using GA along with an adaptive MLP Neural Network classifier, the accuracy of the system is improved and the computational time of the system is reduced.

Selection of classifiers and feature extraction methods has a prime role in achieving best possible classification accuracy in character recognition system. Different classifiers and feature extraction methods for character recognition issues can be resolved through these objectives. In this research work an efficient off-line handwritten character recognition system has been developed by using Support Vector Machine (SVM). Its performance is compared with traditional Multi Layer Perceptron classification techniques. For both the classifiers experiments have been performed using well known standard database acquired from CEDAR (Centre .for Excellence in Document Analysis and Recognition) The experimental results show that the performance of Support Vector Machine (SVM) is better than the Multi Layer Perceptron and other state of art techniques reported in literature.

It is very clearly shown that SVM outperforms MLP for all the three modules of character recognition problem. The performance of the MLP is limited due to the fact that there is no theoretical relationship between the structure of MLP (i.e. number of hidden layers and number of neurons in each layer) and the task of classification. On the other hand the complexities of SVM are independent of size of the dataset. MLP neural network converges on local minima rather than global minima. Principle of risk minimization is the basic difference between SVM and MLP. SVM utilizes Structural Risk Minimization (SRM) principle. Neural network is based on traditional Empirical Risk minimization (ERM) principle. SRM minimizes an upper bound on the expected risk on the other hand ERM reduces the error on the training data. Due to this superior ability, Support Vector Machine gives better generalization performance compared to MLP. Due to the absence of local minima in SVM, it has less testing time compared to MLP. SVM works better than MLP, with standard back-propagation learning algorithm for off-line handwritten character recognition within our dataset. The superior performance of SVM is due to superior generalisation ability of support vector machine in high dimensional space.

Number of useful and important observations has been produced by the research work presented in this thesis. From the observations, some possible useful directions may be implemented to improve the accuracy of character recognition system. A better pre-processing methodology shall be used by using skew correction module. In this research we have not applied method of skew correction on the characters. Different features are having different discriminating abilities for different recognition process. Different varieties of local and global feature extraction techniques could have been applied on the image of a character. The experimental studies presented in this thesis were performed on English alphabet and digits. Considering that our techniques are fairly generic and text independent, their application to other scripts like Chinese, Arabic, and Devanagari is a pertinent and interesting research area. This technique of pattern recognition can also be implemented for recognition of different regional languages. To make the character recognition system more reliable, different classifiers such as SVM, ANN, kNN and Fuzzy Logic can be embedded in a recognition system. Different combination of classifiers will be having different ability to handle discriminant problems. Finally, the mechanism and architecture of the classifiers combination shall also be explored. The character recognizer will also be extended from text to lines to paragraphs and pages.