CHAPTER 8

8. CONCLUSION

The thesis titled "Performance Enhancement of Classifiers for Tamil Writer Identification through Modified SVM Linear Kernel with Parameter Estimation and Deep Learning" portrays the research work carried on Tamil writer identification through machine learning and deep learning approaches.

The goal of this research is to develop discriminative models for writer identification based on Tamil handwritten text images such as character, word and paragraph using pattern classification techniques. Writer Identification models are built using traditional machine learning approach by identifying and extracting the hand crafted features from Tamil handwritings. Deep learning approach aids in building predictive models through self - extraction of features with improved generalization of writer identification. The proposed approaches considerably make the prediction of individuals more perfect and suggest a beneficial solution by combining local and global features of handwritings.

A set of 100 identical paragraphs written by 300 different writers in Tamil language have been collected as corpus and the handwritten texts are converted into images through scanning. The paragraph text is then segmented into words which in turn segmented into characters. 100 text dependent words are chosen for each writer, creating 30000 word samples. Similarly 100 characters are chosen for each writer, creating 30000 character samples. The images in three different corpuses are preprocessed using various image processing techniques. Feature extraction and feature selection processes have been performed and three independent normalized datasets have been generated.

The foremost task in machine learning approach is aggregating the local and global features from the Tamil handwriting text images and to build writer identification models using Support Vector Machines. In learning Support Vector Machine based classifiers three different kernels such as linear, polynomial and RBF have been employed. The three datasets are partitioned into training and testing sets in the ratio of 80% and 20% and three independent classifiers have been built. Performances of the classifiers are analyzed using various measures such as accuracy, precision, recall, F-measure and time taken to build the model. The results of the experiments shows that the SVM with RBF kernel based writer prediction models are able to attain high prediction accuracy than other kernels.

Linear kernel attains very low accuracy compared to other two kernels. But the observation shows that linear kernel performs faster than the other kernels with less computational complexity. The novelty is introduced in linear kernel and the framework of linear kernel is modified using parameter estimation technique. The proposed weighted linear kernel (WLK) is formulated using weighted least square parameter estimation technique and the writer identification models were built using the same three datasets. The use of WLK-SVM is more beneficial in terms of less computational complexity, minimum time, scalability yielding better results than the linear kernel based SVM method. This work motivated to develop another kernel using Bayesian Linear Regression (BLR).

BLR is used to add the regression parameters to the kernel function in order to reduce error rate. The new form of linear kernel BLK defined using BLR is employed in SVM and the writer identification models were built using BLK-SVM for the same three datasets. The practice of this new form of Bayesian Linear Kernel is more favorable and achieved good performance with less computational complexity.

The next model uses another form of linear kernel formulated based on Principal Component Regression. To define Principal Component Kernel, the principal components of the feature matrix are applied and the coefficients of the linear kernel are determined. The PCK with SVM is trained for three datasets and their performances are evaluated for classification of Tamil writing patterns. The new PCK kernel has comparatively higher performance when compared to linear, WLK, BLK kernels. From the experimental results, it is proved that PCK-SVM achieved better performance with minimum time taken and less computational complexity.

Deep learning is an added significant approach exploited for Tamil writer identification. The subsequent work is intended to adopt Convolutional Neural Networks, a kind of deep learning architecture for Tamil Writer identification. The unified framework of CNN enabled feature learning and classification within the deep learning environment thereby reducing the number of tasks required in traditional learning and to build accurate writer prediction model. Comparison between shallow and deep learning approaches is done by measuring up ANN models against CNN models. It is perceived that the predictive accuracy shown by the CNN models is comparatively higher.

Finally a writer identification tool has been developed by integrating CNN based writer identification model with MATLAB GUI for predicting the identity of writer based on their handwritten Tamil text. Here the tool is developed based on offline text dependent writer identification

which can accept a Tamil handwritten text (either a single word text or a paragraph) as input and can predict the writer of the handwriting. The observations made from this research work are summarized as below.

- Global features extracted from texture of image using Gabor filter and GLCM are not enough to predict the writer as the handwriting contain more structural properties. Hence local features are taken into account to build classifiers

- Global features pooled with local features shows better performance to attains an elevated accuracy in writer identification

- Even though linear kernel achieves very less accuracy compared to other kernels in SVM implementation, it is observed that it performs faster than other kernels and also it shows very less computational complexity

- New form of linear kernels defined using parameter estimation techniques, which enriches the performance of SVM models in predictive accuracy

- Traditional supervised classification technique such as SVM yielded desirable accuracy that motivates to extend the research to next level using deep learning

- In ANN fewer number of hidden units produces high training error and increased number of hidden units shows low training error with reduced generalization ability

- CNN is employed and configured with more number of hidden units which confirms less training error and high generalization power in identification of Tamil writer

- CNN is very effective in absorbing shape variations of handwriting since features of handwritten text images are self-extracted by its architecture

- The research contributions made in this thesis are

- Developed real time corpus as there is no benchmark dataset available in Tamil handwriting

- Identified and captured evocative features from character, word and paragraph handwritten text images

- Built efficient discriminative writer identification models using Support Vector Machine, which has strong mathematical foundation

- Performance enhancement of SVM linear kernel using parameter estimation techniques

- Improved generalization of writer identification through a deep learning architecture- convolutional neural network

- Development of Tamil writer identification tool

It is concluded that machine learning techniques are most suitable in providing solution for writer identification problem. The promising results obtained from this research work have encouraged developing new form of linear kernels for SVM using parameter estimation techniques. Finally, the research is directed to deep learning environment using CNN which is very effective in self-extraction of features and modelling Tamil writer identification. This research recommends the significance in forensic expert decision-making, signature verification and biometric authentication. In future, the work can be extended to develop writer identification model in other Indian languages using text independent handwriting images with emerging computational techniques.