CHAPTER-8

CONCLUSION AND FUTURE PERSPECTIVES

Gender Classification and Writer Identification systems are the challenging and stipulating applications based on the offline handwritten text and are widely used for criminal investigations, forensic analysis, forgery detection, and questioned documents. The development of a gender classification system can be based on physiological and behavioural biometric traits. After going through the literature survey, it has been considered that the gender classification system has been successfully developed with the physiological traits such as the face, fingerprints, retina, gate information as compared to the behavioural biometric trait, i.e., handwriting, keystrokes, and so on. On the basis of handwriting trait which is a robust behavioural biometric modality in the Indic script, the development of a gender classification system is a novel and remarkable endeavour in the Gurumukhi script, which is one of the objectives of the proposed work.

In concern with the writer identification system, only a few recognized work has been available for the Gurumukhi script that too with a limited dataset. Therefore the novel task of developing a framework for a gender classification system and achieving an improved accuracy rate for writer identification based on offline handwritten Gurumukhi script has been successfully completed.

This chapter presents the conclusion of the results achieved in this thesis for the gender classification and writer identification system and also summarizing the major contributions and assumptions corresponding to the objectives. In Chapter 1, introduction to biometric systems i.e., physiological and behavioural systems have been presented along with the deep study on the behavioural biometric trait. Handwriting trait or modality with its history, character set, usefulness, and novelty all are discussed in-depth and the difference between male and female handwriting is also presented with necessary images. The handwriting of male and female writers in form of characters, words and sentences have also been represented so as to have a deep understanding of their writing styles. Historical perspective and the character set of the Gurumukhi script are also completely described in chapter 1.

Chapter 2 presents the comprehensive and systematic state-of-the-art work on gender classification and writer identification systems. The literature review on the gender classification system has been presented by covering non-Indic scripts as for the Indic script, the development of the system has not been recognized. Similarly, for the writer identification system, the review of literature has been presented firstly for the Indic scripts, then for the non-Indic scripts, and finally for the multi scripts. The major research gaps with the assumptions and contributions have also been described in this chapter.

In Chapter 3, the data collection process has been explained along with the number of writers, number of samples, and number of classes. In the next section, the framework of the proposed systems has been thoroughly discussed along with the phases with the necessary figures. All the phases such as Pre-processing, feature extraction techniques namely Zoning, Diagonal, Transition, Peak Extent based, and classification phases are elaborated. In Chapter 4, the experiment for the development of gender classification and writer identification system based on the implementation of Zoning, Diagonal, Transition and Peak Extent based feature extraction method has been carried out followed by the hybridization of feature extraction techniques. The experiment successfully revealed high promising accuracy.

Chapter 5 presents a PCA-based gender classification and writer identification system to reduce the dimensionality, computational complexity, and CPU elapsed time and to reveal acceptable accuracy. In Chapter 6, we have discussed the curve fitting-based features and Intersection & Open-end point-based features for the gender classification and writer identification and implementing hybridization of classification techniques with the majority voting scheme. Chapter 7 represents the hybridized approach of feature extraction and classification techniques and yielded maximum accuracy rate.

8.1 MAJOR CONTRIBUTION AND ACHIEVEMENTS OF THE WORK

The work carried out in this thesis has made the following contributions in the field of gender classification and writer identification based on offline handwritten Gurumukhi text.

8.1.1 Generating dataset for Gender Classification and Writer Identification System

For the accomplishment of the objective, we have first generated a corpus for both gender classification and writer identification, in which offline handwritten characters in the Gurumukhi script from 200 writers have been collected. Out of these 200 writers, 100 female writers and 100 male writers participated. Each writer has written 35 primary characters 10 times. So, for gender classification, the female dataset consists of 100×35×10=35000 Gurumukhi characters, and the male dataset consists of 100×35×10=35000 Gurumukhi characters. Similarly, for the writer identification system, the dataset consists of 200×35×10 = 70,000 Gurumukhi characters with 200 writers. By going through the literature reports, it is perceived that no online benchmark dataset in Gurumukhi script has existed publicly for the experimental work of gender classification and writer identification so it is a real, novel, and challenging attempt in this regard.

8.1.2 Hybridization of Feature Extraction Techniques for Gender Classification and Writer Identification System

To develop the gender classification and writer identification systems, the first experiment was based on the implementation of feature extraction techniques followed by the hybridization of feature extraction techniques with initially 150 writers of which 75 were female writers and 75 were male writers. So, after Pre-processing, all the thinned images have been stored and maintained. Feature extraction techniques namely Zoning, Diagonal, Transition, and Peak Extent have been implemented individually and then hybridized approach has been executed followed by the execution of the classification techniques such as K-NN, Decision Tree, Random Forest, and Adaptive Boosting.

For the gender classification, an accuracy rate of 94.6% has been achieved as shown in Table 4.4, and an accuracy of 93.36% has been achieved for writer identification, which is really a remarkable accuracy rate as presented in Table 4.11. Therefore, the objective of implementing feature extraction techniques has been fulfilled successfully. Also, the objective to generate a new feature extraction technique was accomplished based on the hybridization of feature extraction

techniques was implemented and the results achieved for both the systems have been compared in Table 4.9 and Table 4.12.

8.1.3 Principal Component Analysis based Gender Classification and Writer Identification System

Dimensionality reduction is an incredible technique for increasing the efficiency of machine learning techniques. The main target of using PCA is to eliminate irrelevant or extraneous features that are insignificant in the identification process. PCA is based on finding principal components that are orthogonal to each other and hence lead to a reduction in the size of the feature vector. Effect on the CPU elapsed time, accuracy rate, memory requirements, size of the feature vector, are correlated with the dimensionality reduction. In chapter 5, Table 5.1 shows the reduction of feature set after implementing PCA, and Table 5.5 and Table 5.7 show the gender classification accuracy and decrease in CPU elapsed time achieved using the PCA approach. In concern with the classification of gender, the experiment revealed 90.86% accuracy with 8.95 ms CPU elapsed time and regarding the identification of the writer, the experiment revealed an acceptable accuracy rate of 88.13% and a reduction in CPU elapsed time from 17.55 ms to 8.23 ms in Table 5.14. Syntactic analysis on the results attained for gender classification and writer identification have been shown in Table 5.8 and 5.15, respectively.

8.1.4 Curve Fitting and Intersection & Open-End Point based Gender Classification and Writer Identification System

The development of gender classification and writer identification system based on curve fitting and intersection and open-end point-based features has presented in chapter 6. The accuracy rate along with the true positive rate and the false positive rate has been evaluated and analysed. First, the gender classification accuracy rate has been evaluated based on curve fitting features and classification techniques followed by the implementation for writer identification accuracy. Then hybridization of classification technique has been done so as to strengthen the accuracy rate and not to miss any specified instance. Hybridization is accomplished through Majority Voting Scheme which is also called a hard voting scheme. The majority of votes are used for the predicted classification. Results of gender classification and writer identification after using curve fitting-based features and Intersection & open-end points-based

features are depicted in Table 6.9 and 6.10 respectively. The comparative analysis has been shown in Table 6.12 for gender classification and writer identification system. Hybridization of classification techniques facilitates inspirational achievements. Gender classification accuracy of 90.57% and writer identification accuracy of 88.13% has been attained with this method.

8.1.5 Gender Classification and Writer identification based on Majority Voting Scheme

In this experiment, hybridized approach for feature extraction techniques and hybridization of classification techniques using a majority voting scheme is implemented. Zoning, Diagonal, Transition, Peak Extent based features methods are implemented along with hybridization of classification techniques as shown in Table 7.5. Results for gender classification and writer identification without using hybridization and after implementing hybridization are as shown in Table 7.6. So, this is the best experiment that reveals maximum accuracy rates both for gender classification and writer identification for 200 writers with 100 male writers and 100 female writers. An accuracy rate of 94.27% for gender classification and an accuracy rate of 91.93% for writer identification has been realized which is a remarkable and successful achievement with this novel attempt. Comparative analysis has been presented in Table 7.8 for both the gender classification and writer identification system.

8.2 COMPREHENSIVE REPRESENTATION OF EXPERIMENTAL RESULTS

The successful development of a writer identification and gender classification system with a higher accuracy rate is the main objective of this work. Many experimental evaluations have been achieved using different approaches and methodologies. Pre-processing techniques, binarization, normalization and thinning phase, feature extraction, and classification techniques are the main phases in the framework. Principal component analysis for dimensionality reduction has been implemented to experience the increase in accuracy and decrease in CPU elapsed time. Implementing a majority voting scheme to experience hybridization of classification is also implemented. All the results retrieved in the experiments are summarized in Table 8.1

- It has been concluded that for gender classification, maximum accuracy of **94.6% and** writer identification accuracy of **93.93%** has been attained in the commencing stage with 150 writers using hybridization of feature values Zoning, Diagonal, Transition, and Intersection with Adaptive Boosting classifier.
- For gender classification, the maximum accuracy of **94.27%** has been realized with ***200 writers*** with the hybridization of feature extraction Zoning, Diagonal, Transition and Peak Extent and hybridization of classifiers such as ANN, MLP, Decision Trees, Random Forest and Adaptive Boosting. Precision Rate, TPR and FPR are also evaluated to characterize the quality and strength of the experiment.
- For the writer identification system, the maximum accuracy of **91.93%** has been attained reported for **200 *writers*** with the hybridization of feature extraction Zoning, Diagonal, Transition and Peak Extent and hybridization of classifiers such as ANN, MLP, Decision Trees, Random Forest and Adaptive Boosting. Precision Rate, TPR and FPR are also evaluated to characterize the quality and strength of the experiment.

8.3 DISCUSSION

The development of gender classification and writer identification system has been an extremely useful and strenuous application for any kind of investigations and identifications. These systems have numerous utilities in the forensic application, forgery detection, crime investigations in gathering and analyzing handwritten text, etc.

The "Development of Gender Classification and Writer Identification systems for offline handwritten Gurumukhi text" has been successfully implemented and results have been summarized in Table 8.1. The novel work on the gender classification system in Gurumukhi script for 200 writers with an accuracy of 94.27% for 200 writers i.e., 100 female writers and 100 male writers using hybridization of feature extraction techniques and hybridization of classification techniques has been remarkably achieved. For Writer Identification, an improved accuracy rate of 91.93% for 200 writers has been reported with hybridization of feature extraction techniques

with hybridization of classification techniques, which is much better as compared to state-of-the-art work for Gurumukhi script.

Also, it is here to submit that all the research objectives have been successfully, strongly, and well completed that too with a high accuracy rate. Proposal of novel feature extraction algorithm and hybridization of both feature extraction and classification algorithms have also been achieved in the experiments with the boost in the accuracy rate.

A large dataset of 200 writers for writer identification and for gender classification has also been generated with 100 male and 100 female writers. Zoning, Diagonal, Transition, Peak Extent based, Curve fitting, Intersection Open End feature extraction methods have been implemented and the classification algorithms K-NN, SVM, ANN, MLP, RF, and DT have also been carried out with acceptable results, are also executed in various experimental tasks as per our objective.

The experimental work also includes PCA-based dimensionality reduction along with the hybridization of feature values to propose a new feature extractor with an improvement in CPU elapsed time. The majority voting schemes during the hybridization of classification techniques was also implemented to improve the accuracy rates. Various performance evaluation metrics were also evaluated while evaluating the experimental work such as True Positive rate, False positive rate, Root means squared error, Area under curve, and Precision Rate.

Therefore, all the objectives have been successfully realized by using experiments, implemented in chapter 4 to chapter 7, based on different methodologies are reported in Table 8.1 with promising, remarkable, and improved results as compared to state-of-the art work.

8.4 FUTURE PERSPECTIVES

After getting the successful results for the gender classification and writer identification systems, numerous open challenges and novel directions have opened for the handwriting-based research communities. The first is to generate the large size dataset and make it available online. The proposed work can also be extended to the third gender. In addition to the identification of writer and classification of, gender other applications such as left or right-handedness, age, state, nationality, and autopsy

determination can also be very interesting and useful applications based on the Gurumukhi script on which the researchers can move on.

Another novel track for handwriting-based researchers is to implement deep learning models such as AE, CNN, and RNN for classification so that success rates can be reached to new heights of accuracy. This can be made possible by enhancing the dataset to a sufficiently large size as it is the basic requirement of deep models. Pre-processing phase can be made more effective by designing more efficient and optimum tools for scanning, normalization, and thinning.

The proposed work can be extended to the Online Gurumukhi script also. Gender and Writer Identification can be developed for a complete word or a sentence in addition to a character written by the writer i.e., a holistic approach can be implemented. The proposed work can be enhanced to a multi-script rather than a single script. Efficient codes for feature extraction and classification techniques should be developed and implemented so as to retrieve the best-hidden features that can correctly and efficiently identify the writer and the gender. Again for dimensionality reduction, effective codes can be generated for compressing the data. Hybridization of feature extraction and classification techniques can be improved and enhanced efficiently so as to strengthen the results of classification methods.

Development of a gender classification system can be accomplished with other Indic scripts such as Devanagari, Tamil, Bengali, Kannada for which the system has not been developed yet, and to fulfil the research gap too. Also because of the similarity of Gurumukhi characters with other Indic scripts, this work can be expanded with a broad view for the other scripts with a high accuracy rate.

Table 8.1. Summarized View of Experiments Implemented for Gender Classification and Writer Identification System

Proposed Study	**No. of Writers**	**Feature Extraction Methods**	**Classification Methods**	**Hybridization of Features**	**Hybridization of Classification Values**	**PCA based reduction**	**Results Achieved**
Gender Classification System	150	Zoning (F1), Diagonal (F2), Transition (F3), Peak Extent (F4)	K-NN (C1), Decision Tree (C2), Adaptive Boosting (C3), Random Forest(C4)	F1+F2+F3+F4	--	--	Accuracy 94.6%, Precision 94.4%, FPR 2.0%
Writer Identification System	150	Zoning (F1), Diagonal (F2), Transition (F3), Peak Extent (F4)	K-NN (C1), Decision Tree (C2), Adaptive Boosting (C3), Random Forest (C4)	F1+F2+F3+F4	--	--	Accuracy 93.36% Accuracy, TPR 93.23 , FPR 0.39
Gender Classification	200	Zoning (F1), Diagonal (F2), Transition (F3), Peak Extent (F4)	Random Forest, Decision Tree	F1+F2+F3+F4	--	PCA	Accuracy 90.86%, CPU Elapsed Time 8.95ms
Writer Identification System	200	Zoning (F1), Diagonal (F2), Transition (F3), Peak Extent (F4)	Random Forest, Decision Tree	F1+F2+F3+F4	--	PCA	Accuracy 88.13%, CPU Elapsed Time 8.23ms
Gender Classification System	200	Curve fitting based features, Intersection& Open End Point	Linear SVM (C1), K-NN (C2), Decision Tree(C3), Random Forest (C4) Majority voting scheme	--	C1+C2+C3+C4	--	Accuracy 90.57%, TPR 89.47%, FPR 0.36%

Writer Identification System	200	Curve fitting based features, Intersection& Open End Point	Linear SVM (C1), K-NN (C2), Decision Tree (C3), Random Forest (C4) Majority voting scheme	--	C1+C2+C3+C4	--	Accuracy 87.76%, TPR 86.70% FPR 0.41%.
Gender Classification System	200	Zoning (F1), Diagonal (F2), Transition (F3), Peak Extent (F4)	ANN (C1), MLP (C2) Decision Trees (C3), Random Forest (C4), Adaptive Boosting (C5) Majority voting scheme	F1+F2+F3+F4	C1+C2+C3+C4+C5	--	**Maximum Accuracy 94.27%**
Writer Identification System	200	Zoning (F1), Diagonal (F2), Transition (F3), Peak Extent (F4)	ANN(C1), MLP (C2), Decision Tree (C3), Random Forest (C4) Adaptive Boosting (C5) Majority voting scheme	F1+F2+F3+F4	C1+C2+C3+C4+C5	--	**Maximum Accuracy 91.93%**