DEVELOPMENT OF GENDER CLASSIFICATION AND WRITER IDENTIFICATION SYSTEMS FOR OFFLINE HANDWRITTEN GURUMUKHI TEXT

A

THESIS

SUBMITTED TO

MAHARAJA RANJIT SINGH
PUNJAB TECHNICAL UNIVERSITY
BATHINDA (PUNJAB)

IN FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

COMPUTER APPLICATIONS

By

SHAVETA DARGAN
Regd. No: 17414FPE03

Department of Computational Sciences

MAHARAJA RANJIT SINGH PUNJAB

TECHNICALUNIVERSITYBATHINDA (PUNJAB)

2021

CERTIFICATE

I, **ShavetaDargan** hereby certify that the work contained in this thesis entitled "**Development of Gender Classification and Writer Identification Systems For Offline Handwritten Gurumukhi Text"** in fulfillment of requirements for the award of the degree of the **DOCTOR OF PHILOSOPHY** in the Department of Computational Sciences, Maharaja Ranjit Singh Punjab Technical University, Bathinda, is an authentic record of my own work carried under the guidance and supervision of **Dr. Munish Kumar** (Assistant Professor, Department of Computational Sciences, Maharaja Ranjit Singh Punjab Technical University, Bathinda).

The matter presented in this thesis has not been submitted either in part or full to any other University or Institute for the award of any degree.

Dated: (**ShavetaDargan**)

Signature of Candidate

CANDIDATE'S DECLARATION

I hereby certify that the work presented in the thesis, entitled **"Development of Gender Classification and Writer Identification Systems For Offline Handwritten Gurumukhi Text"** in fulfillment of the requirements of the award of the degree of Doctor of Philosophy in the Faculty of Sciences and submitted in Maharaja Ranjit Singh Punjab Technical University, Bathinda is an authentic record of my own work carried out during a period **from January 2018 to June 2021** under the supervision of **Dr. Munish Kumar** (Assistant Professor, Department of Computational Sciences, Maharaja Ranjit Singh Punjab Technical University, Bathinda).

The matter embodied in this thesis has not been submitted by me for the award of any other degree of this or any other University/Institute.

SHAVETA DARGAN

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

(Dr. MUNISH KUMAR)

(Supervisor)
Deptt. of Computational Sciences
Maharaja Ranjit Singh Punjab Technical University,
Bathinda PIN –151001 (INDIA).

The Ph.D. Viva-Voice examination of _____ Research Scholar, has been held on _____.

Sign. of Supervisor **Sign. of External Examiner**

ACKNOWLEDGEMENT

The real strength of attaining a target is through the way of superiority and uncompromising discipline. I wish to express my sincere gratitude to those who have contributed to this thesis and supported me in an endless way in order to make this Ph.D. thesis possible.

First and foremost, I would like to extend my special appreciation and gratitude to my research guide, **Dr. Munish Kumar (Assistant Professor, Department of Computational Sciences, Maharaja Ranjit Singh Punjab Technical University, Bathinda),** for his continuous support, dedicated help, inspiration, encouragement, useful discussions and brainstorming sessions throughout my Ph.D. His personal guidance, constructive criticism, encouragement, and expert advice have been invaluable throughout all the phases of the work. His observations and invaluable feedback helped me to establish the overall direction of the research and to move forward with the investigation in depth.

I am grateful to Dean (R&D), MRSPTU, Bathinda and other DDRC members for their valuable suggestions which have contributed greatly to the improvement of this Ph.D. research work. I also express my gratitude to Prof. Sanjay Bhatnagar (Head, Department of Computational Sciences) and Dr. Amitoj Singh (Assistant Professor, Department of Computational Sciences), Maharaja Ranjit Singh Punjab Technical University, Bathinda for providing me sufficient university resources and timely cooperation in order to carry out the work. I extend my sincere thanks to the writers for helping in generating a dataset by writing Gurumukhi characters.

I also appreciate all administrative, academic, and technical staff from Maharaja Ranjit Singh Punjab Technical University, Bathinda for their timely support and help in their respective roles.

Last but not least, I owe a special thanks to Almighty first then to my husband, my kids, family members and **Director Education, SGPC, Sri Amritsar cum Principal Guru Nanak College, Sri Muktsar Sahib, Dr. Tejinder Kaur Dhaliwal** for their constant encouragement and endless support to accomplish this work.

(SHAVETA DARGAN)

LIST OF FIGURES

Fig. No.	Description	Page No
1.1	Biometric based Recognition Systems	4
1.2	Physiological Biometric Traits	5
1.3	Behavioral Biometric Traits	6
1.4	Hardware Requirement for Offline vs Online handwriting	8
1.5(a)	Sample of Offline handwriting	9
1.5(b)	Sample of Online handwriting	9
1.5(c)	Offline and Online handwriting Samples	9
1.6	Male and Female handwriting	12
1.7(a)	Samples of females handwriting in Gurumukhi script	13
1.7(b)	Samples of males handwriting in Gurumukhi script	13
1.8	Writing Pattern of male and female writers	14
1.9	Word written by male and female writers in Gurumukhi script	15
1.10	Sample of Utterances in the Gurumukhi script	23
3.1	Offline Handwriting Sample of a male writer	53
3.2	Offline Handwriting Sample of a female writer	54
3.3	Framework for Gender Classification and Writer Identification System	56
3.4	Transformation of Digitized to Thinned Image	57
3.5	Sliced Data	57
3.6	Bitmap images of characters	58
3.7	Sample of a thinned Image	58
3.8	Extracting Features with Zoning Method	59
3.9	Hierarchical Pattern of Division	59
3.10	Diagonal Method of Feature Extraction	60
3.11(a)	Transition in Horizontal Direction	61
3.11(b)	Transition in Vertical Direction	61
3.12(a)	Peak Extent Features	61
3.12(b)	Horizontal Peak Extent	61
3.12(c)	Vertical Peak Extent	61
3.13	Curve Fitting Based Method of Feature Extraction	62

3.14	Intersection and Open-End Point Method of Feature Extraction	63
3.15(a)	Conversion of weak to strong classifier	65
3.15(b)	AdaBoost Classification Method	65
3.16	Artificial Neural Network	66
3.17	Decision Tree	67
3.18	K-NN	68
3.19	Random Forest	68
3.20	Multi-Layer Perceptron	69
3.21	Support Vector Machine	70
3.22(a)	Accuracy and Precision	72
3.22(b)	Area Under Curve	72
4.1	Gender Classification Accuracy with hybridization of Feature Extraction Techniques	80
4.2	False Positive Rate for Gender Classification	81
4.3	Precision Rate	82
4.4	Root Mean Square Error	83
4.5	Area Under Curve	85
4.6	Writer Identification Accuracy with hybridization of Feature Extraction Techniques	88
4.7	True Positive Rate for Writer Identification	89
4.8	False Positive Rate for Writer Identification	90
5.1	PCA based representation of multidimensional variables	95
6.1(a)	Fitting of a Parabola curve	108
6.1(b)	Best fit curve	108
6.2	Curve Fitting Features	109
6.3	Intersection and Open-End Point Features	110
6.4	Majority Voting Scheme	112
7.1	Transformation of Digitized Image to Thinned Image	122
7.2	Gender Classification Accuracy without hybridization of feature extraction and classification techniques	124
7.3	Writer Identification Accuracy without hybridization of feature extraction and classification techniques	125
7.4	Gender Classification Accuracy with hybridization of feature extraction and classification techniques	127
7.5	Writer Identification Accuracy with hybridization of feature extraction and classification techniques	128

LIST OF TABLES

Table No	Description	Page No.
1.1	Characteristics of Male and Female handwriting	10
1.2	Character set of the Gurumukhi script	24
2.1	State-of-the-art-work on the Gender Classification System	30
2.2	State-of-the-art on the Writer Identification System based on Non-Indic Scripts	42
2.3	State-of-the-art on the Writer Identification System based on Indic Scripts	46
3.1	Dataset Description for Gender Classification System	52
3.2	Dataset Description for Writer Identification System	53
4.1	Feature values obtained without hybridization of feature extraction techniques	75
4.2	Feature valuesobtained with hybridization of feature extraction techniques	76
4.3	Gender Classification Accuracy with Feature Extraction Techniques	78
4.4	Gender Classification Accuracy with hybridization of Feature Extraction Techniques	79
4.5	False Positive Rate	81
4.6	Precision Rate	82
4.7	Root Mean Square Error	83
4.8	Area Under Curve	84
4.9	Comparative Analysis on Gender Classification Accuracy	85
4.10	Writer Identification Accuracy without hybridization of feature extraction techniques	87
4.11	Writer Identification Accuracy with hybridization of feature extraction techniques	88
4.12	Comparative Analysis on Writer Identification Accuracy	90
5.1	PCA based reduction in feature values	96
5.2	Dataset Description for Gender Classification System	97
5.3	Summarized View of Techniques for Gender Classification System	98
5.4	Experimental Results for Gender Classification without	99

	PCA	
5.5	Experimental Results for Gender Classification with PCA	99
5.6	CPU Elapsed Time and Accuracy Rate for Gender Classification without PCA	100
5.7	CPU Elapsed time and Accuracy Rate for Gender Classification with PCA	100
5.8	Comparative Analysis for Gender Classification System	101
5.9	Description of Dataset for Writer Identification System	102
5.10	Summarized View of Techniques for Writer Identification System	102
5.11	Experimental Results for Writer Identification without PCA	103
5.12	Experimental Results for Writer Identification with PCA	103
5.13	CPU Elapsed time and Accuracy Rate for Writer Identification without PCA	104
5.14	CPU Elapsed time and Accuracy rates for Writer Identification with PCA	104
5.15	Syntactic Analysis for Writer Identification	105
6.1	Dataset Description for Gender Classification System	107
6.2	Gender Classification Accuracy with Curve fitting Features	109
6.3	Writer Identification Accuracy with Curve Fitting Features	109
6.4	Gender Classification Accuracy with Intersection and Open-End Point features	110
6.5	Writer Identification Accuracy with Intersection and Open-End Point features	111
6.6	Gender Classification Accuracy with hybridization of classification techniques and Curve Fitting Features	115
6.7	Writer Identification Accuracy with hybridization of classification techniques and Curve Fitting Features	115
6.8	Gender Classification Accuracy achieved with hybridization of classification techniques and Intersection and Open-End Point based features	116

6.9	Writer Identification Accuracy with hybridization of classification techniques and Intersection and Open-End Point based features	117
6.10	Gender Classification Accuracy with Curve Fitting Based Features and Intersection Open-End Point and hybridization of classification techniques	118
6.11	Writer Identification Accuracy rate with Curve fitting and Intersection and Open-End Point features and hybridization of classification techniques	118
6.12	Results achieved for gender classification and writer identification system without hybridization of classification techniques	119
6.13	Results achieved for gender classification and writer identification system with hybridization of classification techniques	120
7.1	Gender Classification Accuracy	123
7.2	Writer Identification Accuracy	124
7.3	Feature values extracted without hybridization of feature Extraction Techniques	126
7.4	Feature values extracted with hybridization of Feature Extraction Techniques	126
7.5	Gender Classification Accuracy with hybridization of Feature Extraction and Classification Techniques	127
7.6	Writer Identification Accuracy with hybridization of Feature Extraction and Classification Techniques	128
7.7	Comparative Analysis of results achieved for Gender classification and Writer Identification system	129
8.1	Summarized view of experiments implemented for Gender Classification and Writer Identification System	136-137

ABBREVIATIONS

AdaBoost	Adaptive Boosting
AE	Auto-encoder
ALF	Auto Learned Features
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
AUC	Area Under Curve
BPNN	Back propagation Neural Network
BW-LBC	Block Wise-Local Binary Count operator
CNN	Convolution Neural Network
COLD	Cloud of Line Distribution
DAR	Document Analysis and Recognition
DAS	Document Analysis System
DCT	Discrete Cosine Transform
DDC	Distribution Descriptive Curve
DF-ANOVA	Dual factor-Analysis of Variance
DPI	Dots per inch
DT	Decision Tree
FDR	Fisher Discriminant Ratio
FN	False Negative
FN	False Negative
FP	False Positive
FPR	False Positive Rate
GCS	Gender Classification System
GLCM	Gray-Level Co-occurrence Matrices
GMM	Gaussian Mixture Model
HEF	Human Engineered Features
HOG	Histogram of Oriented Gradients
KDA	Kernel Discriminant Analysis
K-NN	K-Nearest Neighbor
KPCA	Kernel Principal Component Analysis
LDA	Linear Discriminant Analysis

LBP	Local Binary Patterns (LBP)
LLE	Local linear embedding
LM	Levenberg Marquardt
LSM	Least Square Method
MDS	Multidimensional Scaling
MLP	Multi-Layer Perceptron
NN	Nearest Neighbor
PC	Principal Components
PCA	Principal Component Analysis
PR	Precision Rate
RBF	Radial Basis function
RF	Random Forest
RMSD	Root Mean Square Deviation
RMSE	Root mean square error
RNN	Recurrent Neural Network
SCI	Science Citation Index
SVM	Support Vector Machine
SFTA	Segmentation-based Fractal Texture Analysis
TN	True Negative
TP	True Positive
TPR	True positive Rate
WIS	Writer Identification System