

CERTIFICATE

I, **Harmandeep Kaur** hereby certify that the work contained in this thesis entitled “**DEVELOPMENT OF OFFLINE HANDWRITTEN GURUMUKHI WORD RECOGNITION SYSTEM FOR POSTAL AUTOMATION**”, in partial fulfillment of requirements for the award of 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.

Date:

(Harmandeep Kaur)

Signature of Candidate

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis, entitled **“DEVELOPMENT OF OFFLINE HANDWRITTEN GURUMUKHI WORD RECOGNITION SYSTEM FOR POSTAL AUTOMATION”** in fulfilment of the requirements of the award of the degree of Doctor of Philosophy in 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 July 2017 to January 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.

Harmandeep Kaur

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(s)

Sign. of External Examiner

ACKNOWLEDGEMENTS

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 the 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 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, 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 writing Gurumukhi words (place names) for the creation of the database without whom this work could not have been possible.

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 my parents Sh. Anteryami Singh and Smt. Manjeet Kaur, and my younger brother Diljot Singh for their constant encouragement and endless support to accomplish this work.

Harmandeep Kaur

LIST OF FIGURES

Figure No.	Title	Page No.
1.1 (a)	Word shape recognition	2
1.1 (b)	Word shape recognition using envelope around the word	2
1.2	Word recognition classification	3
1.3	Schematic diagram of offline handwritten word recognition system	8
1.4	Gurumukhi word partitioned into three zones	15
2.1 (a)	Digitized word image at level L=0	43
2.1 (b)	Digitized word image at level L=1	43
2.1 (c)	Digitized word image at level L=2	43
2.2 (a)	Zones of word image	44
2.2 (b)	Diagonals of Z_{21} zone	44
2.3 (a)	Zones of word image	45
2.3 (b)	Intersection & open end points	45
2.4 (a)	Zoning of bitmap word image	46
2.4 (b)	Horizontal peak extent features	46
2.4 (c)	Vertical peak extent features	46
2.5	Configuration of k-NN classifier	48

2.6	Configuration of SVM classifier	49
2.7	Decision Tree	50
2.8	Configuration of Random Forest classifier	51
3.1	Offline handwritten Gurumukhi words	65
4.1	Working of AdaBoost Methodology	73
4.2	System evaluation based on Accuracy	78
4.3	System evaluation based on FAR	79
4.4	System evaluation based on FRR	80
5.1 (a)	Horizontal boundary extent feature extraction	86
5.1 (b)	Vertical boundary extent feature extraction	86
5.2	Performance comparison of feature selection techniques	91
6.1	Working of Bagging methodology	95
7.1	Working of XGBoost technique	104
7.2	Area Under Curve (AUC)	105
7.3	System evaluation using zoning features	107
7.4	System evaluation using diagonal features	108
7.5	System evaluation using intersection & open end points features	109
7.6	System evaluation using peak extent features	110

LIST OF TABLES

Table No.	Title	Page No.
1.1	Comparison between Online and Offline Handwritten Word Recognition	5
1.2	Gurumukhi characters with pronunciation	13
1.3	Gurumukhi Vowels	14
1.4	Subscript letters	14
1.5	Auxiliary signs	14
2.1	Recognition results for Indic Scripts	53
2.2	Recognition results for Non-Indic Scripts	57
2.3	Recognition results for Bi-Scripts and Multi-Scripts	61
3.1	Database description	65
3.2	Benchmark dataset description	66
3.3	Offline handwritten Gurumukhi words	67
3.4	Thinned images corresponding to digitized images of handwritten Gurumukhi words	68
4.1	System evaluation based on Accuracy	78
4.2	System evaluation based on FAR	79
4.3	System evaluation based on FRR	80
4.4	Comparison of the present approach with state-of-the-art approaches	81

5.1	Performance comparison of feature selection techniques	90
5.2	Performance (TPR) comparison of feature selection techniques based on 5-fold cross-validation	91
6.1	Recognition results based on the considered features and classifiers without Bagging methodology	96
6.2	Recognition results based on Bagging methodology	98
6.3	Comparative analysis of the present work and state-of-the-art work	100
7.1	Dataset Partitioning Strategies	106
7.2	System evaluation using zoning features	107
7.3	System evaluation using diagonal features	108
7.4	System evaluation using intersection & open end points features	109
7.5	System evaluation using peak extent features	110
7.6	Comparison of the present approach with state-of-the-art approaches	111
7.7	Comparison of the present approach with state-of-the-art Gurumukhi character recognition approaches	113

LIST OF ABBREVIATIONS

AdaBoost	Adaptive Boosting
AFDM	Adversarial Feature Deformation Module
AUC	Area Under Curve
Bagging	Bootstrap Agregating
CAR	Character Accurate Rate
CBA	Consistency Based Analysis
CFS	Correlation Feature Set
CM	Chebyshev Moments
CNN	Convolutional Neural Network
CNN-TL	Convolutional Neural Network- Transfer Learning
CON	Consistency-Based
CRNN	Convolutional Recurrent Neural Network
CSA	Chi Squared Attribute
DCNN	Deep Convolutional Neural Network
DCT	Discrete Cosine Transform
DDD	Directional Distance Distribution
DF	Fourier Descriptors
DHBN	Dynamic Hierarchical Bayesian Network
DPI	Dots Per Inch

ELM	Extreme Learning Machine
FAR	False Acceptance Rate
FN	False Negative
FP	False Positive
FPR	False Positive Rate
FRR	False Rejection Rate
GA	Genetic Algorithm
GBM	Gradient Boosting Machine
GSC	Gradient Structural Concavity
HS	Harmony Search
HMM	Hidden Markov Model
HNC	Hierarchical Network of Concept
HOG	Histogram of Oriented Gradient
HWR	Handwritten Word Recognition
k -NN	k -Nearest Neighbor
LPP	Locality Preserving Projections
LSTM	Long Short Term Memory
MA	Memetic Algorithm
MLP	Multilayer Perceptron
MLP-NN	Multilayer Perceptron-Neural Network
MQDF	Modified Quadratic Discriminant Function

NB	Naive Bayes
NN	Neural Network
NN-MLPs	Neural Network-Multilayer Perceptrons
OCR	Optical Character Recognition
PAW	Part-of-Arabic-Word
PCA	Principal Component Analysis
PHOG	Pyramid Histogram of Oriented Gradient
PSO	Particle Swarm Optimization
RBF	Radial Basis Function
RMSD	Root Mean Squared Deviation
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristics
SCF	Statistical and Contour based Features
SFFS	Sequential Forward Floating Search
SGCSL	Statistical Geometric Components of Straight Lines
SMO	Sequential Minimal Optimization
SOFM	Self Organizing Feature Map
SVM	Support Vector Machine
TAN	Tree Augmented Naive bays network
TN	True Negative

TP	True Positive
TPR	True Positive Rate
VH-HMM	Horizontal and Vertical Hidden Markov Model
WER	Word Error Rate
XGBoost	eXtreme Gradient Boosting