

# CONTENTS

---

<b>CERTIFICATE</b>	<b>ii</b>
<b>CANDIDATE’S DECLARATION</b>	<b>iii</b>
<b>ABSTRACT</b>	<b>iv-vii</b>
<b>ACKNOWLEDGEMENTS</b>	<b>ix</b>
<b>CONTENTS</b>	<b>x-xvi</b>
<b>LIST OF FIGURES</b>	<b>xvii-xviii</b>
<b>LIST OF TABLES</b>	<b>xix-xx</b>
<b>LIST OF PUBLICATIONS</b>	<b>xxi</b>
<b>LIST OF ABBREVIATIONS</b>	<b>xxii-xxv</b>
<b>CHAPTER 1: INTRODUCTION</b>	<b>1-17</b>
1.1 Background of Word Recognition System	2
1.1.1 Printed Word Recognition	3
1.1.1.1 Good Quality Printed Word Recognition	3
1.1.1.2 Degraded Printed Word Recognition	3
1.1.2 Handwritten Word Recognition (HWR)	4
1.1.2.1 Online Handwritten Word Recognition	4
1.1.2.2 Offline Handwritten Word Recognition	4
1.1.3 Approaches to Handwritten Word Recognition	5

1.1.3.1	Analytical Approach	6
1.1.3.2	Holistic Approach	6
1.1.4	Pros of Word Recognition	7
1.1.5	Cons of Word Recognition	7
1.2	Phases of an Offline Handwritten Word Recognition System	8
1.2.1	Digitization	9
1.2.2	Pre-processing	9
1.2.3	Segmentation	9
1.2.4	Feature Extraction	10
1.2.5	Feature Selection	10
1.2.6	Classification	11
1.2.7	Post-processing	11
1.3	Applications of an Offline Handwritten Word Recognition System	11
1.4	Overview of Gurumukhi Script	12
1.5	Objectives of the Work	15
1.6	Assumptions	15
1.7	Major Contributions and Achievements	16
1.8	Organization of the Thesis	16
<b>CHAPTER 2: REVIEW OF LITERATURE</b>		<b>18-63</b>

2.1	Non-Indic Scripts	18
2.1.1	Arabic	18
2.1.2	Chinese	22
2.1.3	Dutch	23
2.1.4	Japanese	23
2.1.5	Latin/Roman	24
2.1.6	Mongolian	25
2.1.7	Persian/Farsi	26
2.1.8	Thai	27
2.1.9	Uyghur	27
2.2	Indic Scripts	28
2.2.1	Bangla	28
2.2.2	Devanagari	31
2.2.3	Gujarati	32
2.2.4	Gurumukhi	33
2.2.5	Kannada	35
2.2.6	Malayalam	36
2.2.7	Oriya	37
2.2.8	Tamil	37

2.3	Bi-Scripts	38
2.4	Multi-Scripts	40
2.5	Algorithms used in this work at distinct phases of recognition system	41
2.5.1	Digitization	42
2.5.2	Pre-processing	42
2.5.3	Feature Extraction	42
2.5.3.1	Zoning features	42
2.5.3.2	Centroid features	43
2.5.3.3	Diagonal features	43
2.5.3.4	Intersection & open-end points features	44
2.5.3.5	Peak extent features	45
2.5.4	Classification	47
2.5.4.1	k-NN classifier	47
2.5.4.2	SVM classifier	49
2.5.4.3	Decision Tree classifier	50
2.5.4.4	Random Forest classifier	51
2.6	Recognition accuracy attained for Indic and Non-Indic Scripts	52
2.7	Research Gaps	62
2.8	Chapter Summary	63

**CHAPTER 3: DATA COLLECTION, DIGITIZATION AND PRE-PROCESSING** **64-69**

3.1	Data Collection	64
3.2	Digitization	66
3.3	Pre-processing	66
3.4	Chapter Summary	69

**CHAPTER 4: OFFLINE HANDWRITTEN GURUMUKHI WORD RECOGNITION SYSTEM BASED ON HOLISTIC APPROACH AND ADABOOST METHODOLOGY** **70-84**

4.1	Holistic Approach	70
4.2	Feature Extraction Techniques	71
4.2.1	Zoning features	71
4.2.2	Diagonal features	71
4.2.3	Intersection & open-end points features	72
4.3	AdaBoost Methodology	72
4.4	Experimental Results	74
4.4.1	System performance based on k-NN classifier	76
4.4.2	System performance based on RBF-SVM classifier	76
4.4.3	System performance based on Random Forest classifier	76
4.4.4	System performance based on Majority Voting scheme	77

4.4.5	System performance based on AdaBoost Methodology	77
4.5	Comparison with the existing approaches and Syntactic Analysis	81
4.6	Chapter Summary	83
<b>CHAPTER 5: FEATURE SELECTION TECHNIQUES FOR OFFLINE HANDWRITTEN GURUMUKHI WORD RECOGNITION</b>		<b>85-92</b>
5.1	Boundary Extent Feature Extraction Technique	85
5.2	Feature Selection Techniques	86
5.2.1	CBA	86
5.2.2	CFS	87
5.2.3	CSA	87
5.2.4	PCA	87
5.3	Experimental Results	88
5.4	Analysis based on Experimental Results	92
5.5	Chapter Summary	92
<b>CHAPTER 6: BAGGING METHODOLOGY FOR OFFLINE HANDWRITTEN GURUMUKHI WORD RECOGNITION</b>		<b>93-102</b>
6.1	Feature Extraction Techniques	93
6.1.1	Zoning features	93
6.1.2	Centroid features	94

6.1.3	Diagonal features	94
6.1.4	Peak extent features	94
6.2	Bagging Methodology	94
6.3	Experimental Results	96
6.4	Comparison with the existing approaches and Syntactic Analysis	100
6.5	Chapter Summary	102
<b>CHAPTER 7: RECOGNITION OF OFFLINE HANDWRITTEN GURUMUKHI WORDS USING EXTREME GRADIENT BOOSTING (XGBOOST) TECHNIQUE</b>		<b>103-115</b>
7.1	XGBoost Technique	103
7.2	Evaluation Measures	105
7.3	Experimental Results	105
7.3.1	System performance based on zoning features	106
7.3.2	System performance based on diagonal features	107
7.3.3	System performance based on intersection & open end points features	108
7.3.4	System performance based on peak extent features	109
7.4	Comparison with the existing approaches and Syntactic Analysis	110
7.5	Chapter Summary	115
<b>CHAPTER 8: CONCLUSIONS AND FUTURE DIRECTIONS</b>		<b>116-120</b>

8.1	Brief Contribution of the Work	117
8.2	Conclusions	119
8.3	Future Directions	120
	<b>REFERENCES</b>	<b>121-141</b>