

CONTENTS

Acknowledgement	i
List of Tables	iii
List of Figures	v
List of Abbreviations	viii
Abstract	x
1 INTRODUCTION	1
1.1. Data Mining	1
1.2. Writer Identification	8
1.2.1. Writer Identification as a Behavioural Biometric	9
1.2.2. Writer Identification in Forensics	10
1.2.3. Writer Identification vs. Handwriting Recognition	11
1.2.4. Text-Dependent vs. Text-independent Methods	13
1.2.5. Automatic Writer Identification Framework	13
1.3. Need for Automatic Writer Identification	15
1.4. Review of Literature	16
1.5. Objectives of Research	24
1.6. Organization of the Thesis	25
2 MACHINE LEARNING	27
2.1. Supervised Classification Algorithm	28
2.2. Support Vector Machine	29
2.2.1. Geometrical Interpretation of SVM	29
2.2.2. SVM Formulation - Hard Margin	34
2.2.3. SVM Formulation with Soft Margin - L_1 Norm	38
2.2.4. L_2 Norm Linear SVM	43
2.2.5. Non-linear SVM and Kernel Trick	46

2.2.6. SVM Formulation of Non-Linear Kernels with Soft Margin - L_1 Norm	49
2.2.7. Multi-Class Support Vector Machines	51
2.3. Deep Learning	55
2.3.1. Artificial Neural Network	56
2.3.2. Deep Learning Architecture	63
2.4. Summary	69
3 PROBLEM MODELING	70
3.1. Corpus Preparation	71
3.2. Pre-processing	74
3.2.1. Character and Word Text Image Pre-processing	75
3.2.2. Paragraph Text Image Pre-processing	78
3.3. Feature Extraction	81
3.3.1. Feature Extraction from Character / Word Text Images	81
3.3.2. Feature Extraction from Paragraph Text Images	89
3.4. Feature Selection	103
3.5. Datasets	104
3.6. Training and Testing	104
3.7. Summary	107
4 WRITER IDENTIFICATION MODEL THROUGH SVM WITH EXISTING KERNELS	109
4.1 Model I - Writer Identification Model using Support Vector Machine (SVM)	109
4.2 Summary	121
5 WRITER IDENTIFICATION MODEL THROUGH SVM WITH MODIFIED LINEAR KERNELS	123
5.1. Model II - Writer Identification Model through SVM with Weighted Least Kernel (WLK)	123

5.2. Model III - Writer Identification Model through SVM with Bayesian Linear Kernel (BLK)	129
5.3. Model IV - Writer Identification Model through SVM with Principal Component Kernel (PCK)	134
5.4. Summary	141
6 WRITER IDENTIFICATION MODEL THROUGH DEEP LEARNING	143
6.1. Model V - Writer Identification Model through Convolutional Neural Networks (CNN)	143
6.2 Comparison of Models Based on ANN and CNN	146
6.3. Summary	152
7 WRITER IDENTIFICATION TOOL	154
7.1. Development of Tool	154
7.2. Summary	157
8 CONCLUSION	158
REFERENCES	162
LIST OF PUBLICATIONS	169
Appendix	
A. Sample Tamil Handwritings (Paragraphs)	
B. Sample Datasets	
C. Sample Code	
D. Sample Screen Shots – Writer Identification Tool	