APPLICATION OF GEOMETRIC MORPHOMETRIC (GMM) APPROACH FOR AUTHOR'S ETHNICITY DISCRIMINATION USING THE HANDWRITTEN NUMERAL CHARACTERS

WAN NURUL SYAFAWANI BINTI WAN MOHD TAUFEK

UNIVERSITI SAINS MALAYSIA

2024

APPLICATION OF GEOMETRIC MORPHOMETRIC (GMM) APPROACH FOR AUTHOR'S ETHNICITY DISCRIMINATION USING THE HANDWRITTEN NUMERAL CHARACTERS

by

WAN NURUL SYAFAWANI BINTI WAN MOHD TAUFEK

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

March 2024

ACKNOWLEDGEMENT

I would like to express my gratitude to Allah SWT, for His many blessings that have made it possible for me to complete my Doctor of Philosophy journey. It is indeed a long and lonely journey, but if all difficulties were known at the outset, I or most of us would probably never start out at all. I would like to take this opportunity to say my heartfelt gratitude and thanks to my supervisor, Dr Dzulkiflee Ismail, and my co-supervisor, Dr Helmi Mohd Hadi Pritam, who gave me the golden opportunity to conduct this wonderful research project entitled 'Application of Geometric Morphometric (GMM) Approach for Author's Ethnicity Discrimination Using the handwritten numeral characters'. I want to extend my deepest gratitude to all my forensic science lecturers, especially to Dr Wan Nur Syuhaila, who has been with me constantly from the Day-1 until the day this thesis was born. I am extremely grateful to Universiti Sains Malaysia (USM), Health Campus for providing research facilities as well as the equipment and spaces which we needed. I would like to recognise the invaluable assistance of Main Campus, Engineering Campus and Health Campus staff as well as the USM Student Representative Council (MPP), and the 390 participants, for their contribution in this study. Besides, I would like to pay my special regards to my parents; Wan Mohd Taufek bin Hj Wan Ab Hamid and Harlina binti Othman, my siblings and my cats for their unwavering physical, financial and mental support throughout my postgraduate journey. At last, but not in least, to all my friends (you know who you are), I want to express my heartfelt appreciation to all of you who has contributed directly or indirectly in this research. Thank you for all your encouraging words and your time for helping me with my research project.

TABLE OF CONTENTS

ACKN	NOWLEE	DGEMENT	ii
TABL	E OF CC	ONTENTS	iii
LIST	OF TABI	LES	vii
LIST	OF FIGU	JRES	viii
LIST	OF SYM	BOLS	xvii
LIST	OF ABBI	REVIATIONS	xviii
LIST	OF APPE	ENDICES	xxii
ABST	RAK		xxiii
ABST	RACT		XXV
CHAI	PTER 1	INTRODUCTION	1
1.1	Chapter	Overview	1
1.2	Study Ba	ackground	1
1.3	Numeral	handwriting	4
1.4	Geometr	ic Morphometric (GMM)	7
1.5	Problem	Statement & Study Rationale	
1.6	Research	a Question(s)	
1.7	Significa	nce of the Study	
1.8	Objective	es	
	1.8.1	General	
	1.8.2	Specific Objectives	
CHAI	PTER 2	LITERATURE REVIEW	14
2.1	Chapter	Overview	14
2.2	The Unic	queness of Handwriting	14
	2.2.1	Handwriting as an Evidence	

2.3	Factors	Influenced Handwriting	
	2.3.1	Ethnicity Factor Influenced on Handwriting	
2.4	History	and Evolution of the Handwritten Numeral Characters	
	2.4.1	Studies on Handwritten Numeral Characters	
2.5	Geomet	ric Morphometric (GMM) Approach	
	2.5.1	Studies of Geometric Morphometric (GMM) Approach56	
2.6	Statistic	al Analysis in the Geometric Morphometric (GMM) Approach 64	
	2.6.1	Principal Component Analysis (PCA)	
	2.6.2	Procrustes Analysis of Variance (ANOVA)70	
	2.6.3	Discriminant Function Analysis (DFA)78	
СНА	PTER 3	METHODOLOGY 82	
3.1	Chapter	Overview	
3.2	Study A	Study Area	
3.3	Specime	cimen Size Estimation	
3.4	Study Po	opulation and Materials	
	3.4.1	Subject Criteria	
		3.4.1(a) Inclusion Criteria	
		3.4.1(b) Exclusion Criteria	
3.5	Samplin	g Strategy	
3.6	The Har	ndwritten Numeral Characters Discrimination Process	
	3.6.1	Class Characteristics and Landmarking Points Extraction	
	3.6.2	Five Significant Stages for the Handwritten Numeral Character Discrimination Process	
СНА	PTER 4	RESULTS AND DISCUSSIONS 100	
4.1	Chapter	Overview 100	
4.2	Principal Components Analysis (PCA) for Visualisation and Discrimination100		
	4.2.1	The Handwritten Numeral Character of 0104	

	4.2.1(a)	Malay vs Chinese Authors106
	4.2.1(b)	Malay vs Indian Authors108
	4.2.1(c)	Chinese vs Indian Authors110
4.2.2	The Hand	dwritten Numeral Character of 1113
	4.2.2(a)	Malay vs Chinese Authors114
	4.2.2(b)	Malay vs Indian Authors117
	4.2.2(c)	Chinese vs Indian Authors119
4.2.3	The Hand	dwritten Numeral Character of 2122
	4.2.3(a)	Malay vs Chinese Authors
	4.2.3(b)	Malay vs Indian Authors126
	4.2.3(c)	Chinese vs Indian Authors
4.2.4	The Hand	dwritten Numeral Character of 3131
	4.2.4(a)	Malay vs Chinese Authors
	4.2.4(b)	Malay vs Indian Authors135
	4.2.4(c)	Chinese vs Indian Authors
4.2.5	The Hand	dwritten Numeral Character of 4140
	4.2.5(a)	Malay vs Chinese Authors
	4.2.5(b)	Malay vs Indian Authors144
	4.2.5(c)	Chinese vs Indian Authors147
4.2.6	The Hand	dwritten Numeral Character of 5150
	4.2.6(a)	Malay vs Chinese Authors152
	4.2.6(b)	Malay vs Indian Authors154
	4.2.6(c)	Chinese vs Indian Authors157
4.2.7	The Hand	dwritten Numeral Character of 6160
	4.2.7(a)	Malay vs Chinese Authors161
	4.2.7(b)	Malay vs Indian Authors163
	4.2.7(c)	Chinese vs Indian Authors166

	4.2.8	The Handwritten Numeral Character of 7168
		4.2.8(a) Malay vs Chinese Authors
		4.2.8(b) Malay vs Indian Authors
		4.2.8(c) Chinese vs Indian Authors
	4.2.9	The Handwritten Numeral Character of 8177
		4.2.9(a) Malay vs Chinese Authors
		4.2.9(b) Malay vs Indian Authors
		4.2.9(c) Chinese vs Indian Authors
	4.2.10	The Handwritten Numeral Character of 9186
		4.2.10(a) Malay vs Chinese Authors
		4.2.10(b) Malay vs Indian Authors
		4.2.10(c) Chinese vs Indian Authors
4.3	Procrust	es Analysis of Variance (ANOVA) 199
Discr	riminant Fu	nction Analysis (DFA) 205
СНА	PTER 5	CONCLUSION
5.1	Conclus	on
5.2	Limitati	ons of the study 217
5.3	Recomm	endations for Future Works 222
REF	ERENCE	5
APPENDICES		
LIST OF CONFERENCES AND WORKSHOP		
PUBLICATIONS		

vi

LIST OF TABLES

Table 3.1	Specimen size calculation from Raosoft Inc. software	83
Table 3.2	The class characteristics of the handwritten numeral character 0 to 9 identified from the 390 participants	90
Table 3.3	Definition of landmark configurations of each the handwritten numeral characters	92
Table 3.4	Continued	93
Table 3.5	Continued	94
Table 3.6	Continued	95
Table 4.1	Results of Procrustes ANOVA statistics tests for compared groups between Malay vs Chinese authors describing shape variables	202

LIST OF FIGURES

Page

Figure 1.1	The handwritten numeral characters of eight different languages (Bulliet et al., 2019)
Figure 1.2	Handwritten and printed numerals since the late 15th century (Ifrah and Bair, 1987)
Figure 2.1	Number of sense tell us in a glance if a collection of element contains one, two, three or four or more (Ifrah and Bair, 1987)37
Figure 2.2	The first principle primitive people use to represent a whole number (Ifrah and Bair, 1987)
Figure 2.3	Unlimited number of symbols for keeping record of the count (Ifrah, 2000)
Figure 3.1	The snapshot of Raosoft Inc. calculate the specimen size from the Malaysia population size in 2019
Figure 3.2	The proposed framework methodologies of this study
Figure 4.1	Example of raw coordinates in tps format before import into MorphoJ software
Figure 4.2	The eigenvalue plot explaining the total variance of 97.96% between Malay vs Chinese authors for the handwritten numeral character of 0
Figure 4.3	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Malay vs Chinese authors for the handwritten numeral character of 0
Figure 4.4	The 3D PCA scatterplot PCI vs PC2 vs PC3 of the handwritten numeral character of 0 between Malay vs Chinese authors
Figure 4.5	The eigenvalue plot explaining the total variance of 96.89% between Malay vs Indian authors for the handwritten numeral character of 0

Figure 4.6	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Malay vs Indian authors for the handwritten numeral character of 0
Figure 4.7	The 3D PCA scatterplot PCI vs PC2 vs PC3 of the handwritten numeral character of 0 between Malay vs Indian authors110
Figure 4.8	The eigenvalue plot explaining the total variance of 97.14% between Chinese vs Indian authors for the handwritten numeral character of 0
Figure 4.9	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Chinese vs Indian authors for the handwritten numeral character of 0
Figure 4.10	The 3D PCA scatterplot PCI vs PC2 vs PC3 of the handwritten numeral character of 0 between Chinese vs Indian authors
Figure 4.11	The eigenvalue plot explaining the total variance of 83.49% between Malay vs Chinese authors for the handwritten numeral character of 1
Figure 4.12	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Malay vs Chinese authors for the handwritten numeral character of 1
Figure 4.13	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Malay vs Chinese authors for the handwritten numeral character of 1
Figure 4.14	The eigenvalue plot explaining the total variance of 83.63% between Malay vs Indian authors for the handwritten numeral character of 1
Figure 4.15	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Malay vs Indian authors for the handwritten numeral character of 1
Figure 4.16	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Malay vs Indian authors for the handwritten numeral character of 1119

Figure 4.17	The eigenvalue plot explaining the total variance of 85.07% between Chinese vs Indian authors for the handwritten numeral character of 1
Figure 4.18	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Chinese vs Indian authors for the handwritten numeral character of 1
Figure 4.19	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Chinese vs Indian authors for the handwritten numeral character of 1121
Figure 4.20	The eigenvalue plot explaining the total variance of 71.74% between Malay vs Chinese authors for the handwritten numeral character of 2
Figure 4.21	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Malay vs Chinese authors for the handwritten numeral character of 2
Figure 4.22	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Malay vs Chinese authors for the handwritten numeral character of 2
Figure 4.23	The eigenvalue plot explaining the total variance of 72.59% between Malay vs Indian authors for the handwritten numeral character of 2
Figure 4.24	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Malay vs Indian authors for the handwritten numeral character of 2
Figure 4.25	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Malay vs Indian authors for the handwritten numeral character of 2128
Figure 4.26	The eigenvalue plot explaining the total variance of 73.81% between Chinese vs Indian authors for the handwritten numeral character of 2
Figure 4.27	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Chinese vs Indian authors for the handwritten numeral character of 2

Figure 4.28	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Chinese vs Indian authors for the handwritten numeral character of 2131
Figure 4.29	The eigenvalue plot explaining the total variance of 70.78% between Malay vs Chinese authors for the handwritten numeral character of 3
Figure 4.30	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Malay vs Chinese authors for the handwritten numeral character of 3
Figure 4.31	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Malay vs Chinese authors for the handwritten numeral character of 3
Figure 4.32	The eigenvalue plot explaining the total variance of 68.92% between Malay vs Indian authors for the handwritten numeral character of 3
Figure 4.33	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Malay vs Indian authors for the handwritten numeral character of 3
Figure 4.34	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Malay vs Indian authors for the handwritten numeral character of 3137
Figure 4.35	The eigenvalue plot explaining of the total variance of 71.70% between Chinese vs Indian authors for the handwritten numeral character of 3
Figure 4.36	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Chinese vs Indian authors for the handwritten numeral character of 3
Figure 4.37	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Chinese vs Indian authors for the handwritten numeral character of 3140
Figure 4.38	The eigenvalue plot explaining the total variance of 62.71% between Malay vs Chinese authors for the handwritten numeral character of 4

Figure 4.39	The 2D PCA scatterplot of the PC1 vs PC2 clustering between Malay vs Chinese authors for the handwritten numeral character of 4
Figure 4.40	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Malay vs Chinese authors for the handwritten numeral character of 4
Figure 4.41	The eigenvalue plot explaining the total variance of 65.38% between Malay vs Indian authors for the handwritten numeral character of 4
Figure 4.42	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Malay vs Indian authors for the handwritten numeral character of 4
Figure 4.43	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Malay vs Indian authors for the handwritten numeral character of 4147
Figure 4.44	The eigenvalue plot explaining the total variance of 63.30% between Chinese vs Indian authors for the handwritten numeral character of 4
Figure 4.45	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Chinese vs Indian authors for the handwritten numeral character of 4
Figure 4.46	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Chinese vs Indian authors for the handwritten numeral character of 4149
Figure 4.47	The eigenvalue plot explaining the total variance of 63.51% between Malay vs Chinese authors for the handwritten numeral character of 5
Figure 4.48	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Malay vs Chinese authors for the handwritten numeral character of 5
Figure 4.49	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Malay vs Chinese authors for the handwritten numeral character of 5

Figure 4.50	The eigenvalue plot explaining the total variance of 65.75% between Malay vs Indian authors for the handwritten numeral character of 5
Figure 4.51	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Malay vs Indian authors for the handwritten numeral character of 5
Figure 4.52	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Malay vs Indian authors for the handwritten numeral character of 5157
Figure 4.53	The eigenvalue plot explaining the total variance of 62.71% between Chinese vs Indian authors for the handwritten numeral character of 5
Figure 4.54	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Chinese vs Indian authors for the handwritten numeral character of 5
Figure 4.55	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Chinese vs Indian authors for the handwritten numeral character of 5159
Figure 4.56	The eigenvalue plot explaining the total variance of 77.65% between Malay vs Chinese authors for the handwritten numeral character of 6
Figure 4.57	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Malay vs Chinese authors for the handwritten numeral character of 6
Figure 4.58	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Malay vs Chinese authors for the handwritten numeral character of 6
Figure 4.59	The eigenvalue plot explaining the total variance of 74.36% between Malay vs Indian authors for the handwritten numeral character of 6
Figure 4.60	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Malay vs Indian authors for the handwritten numeral character of 6

Figure 4.61	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Malay vs Indian authors for the handwritten numeral character of 6165
Figure 4.62	The eigenvalue plot explaining the total variance of 77.50% between Chinese vs Indian authors for the handwritten numeral character of 6
Figure 4.63	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Chinese vs Indian authors for the handwritten numeral character of 6
Figure 4.64	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Chinese vs Indian authors for the handwritten numeral character of 6168
Figure 4.65	The eigenvalue plot explaining the total variance of 93.03% between Malay vs Chinese authors for the handwritten numeral character of 7
Figure 4.66	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Malay vs Chinese authors for the handwritten numeral character of 7
Figure 4.67	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Malay vs Chinese authors for the handwritten numeral character of 7
Figure 4.68	The eigenvalue plot explaining the total variance of 93.29% between Malay vs Indian authors for the handwritten numeral character of 7
Figure 4.69	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Malay vs Indian authors for the handwritten numeral character of 7
Figure 4.70	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Malay vs Indian authors for the handwritten numeral character of 7174
Figure 4.71	The eigenvalue plot explaining the total variance of 92.74% between Chinese vs Indian authors for the handwritten numeral character of 7

Figure 4.72	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Chinese vs Indian authors for the handwritten numeral character of 7
Figure 4.73	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Chinese vs Indian authors for the handwritten numeral character of 7
Figure 4.74	The eigenvalue plot explaining the total variance of 64.18% between Malay vs Chinese authors for the handwritten numeral character of 8
Figure 4.75	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Malay vs Chinese authors for the handwritten numeral character of 8
Figure 4.76	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Malay vs Chinese authors for the handwritten numeral character of 8
Figure 4.77	The eigenvalue plot explaining the total variance of 64.80% between Malay vs Indian authors for the handwritten numeral character of 8
Figure 4.78	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Malay vs Indian authors for the handwritten numeral character of 8
Figure 4.79	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Malay vs Indian authors for the handwritten numeral character of 8183
Figure 4.80	The eigenvalue plot explaining the total variance of 63.87% between Chinese vs Indian authors for the handwritten numeral character of 8
Figure 4.81	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Chinese vs Indian authors for the handwritten numeral character of 8
Figure 4.82	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Chinese vs Indian authors for the handwritten numeral character of 8

Figure 4.83	The eigenvalue plot explaining the total variance of 69.43% between Malay vs Chinese authors for the handwritten numeral character of 9
Figure 4.84	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Malay vs Chinese authors for the handwritten numeral character of 9
Figure 4.85	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Malay vs Chinese authors for the handwritten numeral character of 9
Figure 4.86	The eigenvalue plot explaining the total variance of 67.19% between Malay vs Indian authors for the handwritten numeral character of 9
Figure 4.87	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Malay vs Indian authors for the handwritten numeral character of 9
Figure 4.88	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Malay vs Indian authors for the handwritten numeral character of 9
Figure 4.89	The eigenvalue plot explaining the total variance of 70.34% between Chinese vs Indian authors for the handwritten numeral character of 9
Figure 4.90	The 2D PCA scatterplot of the PC1 vs PC2 with clustering between Chinese vs Indian authors for the handwritten numeral character of 9
Figure 4.91	The 3D PCA scatterplot PCI vs PC2 vs PC3 between Chinese vs Indian authors for the handwritten numeral character of 9 194

LIST OF SYMBOLS

H _A	alternative hypothesis
\mathbb{R}^2	coefficient of determination
df	degrees of freedom
=	equal to
gsm	grams per square meter
mm	Millimetre
Ho	null hypothesis
%	percentage
+	Plus
q	proportion of non-response rate
R	Repeatability
n	specimen size required
N_1	specimen size to be recruited
CD	400
CLXXXVII	100+50+10+10+10+5+1+1
IV	4
MDCXXVI	1000+500+100+10+10+5+1

LIST OF ABBREVIATIONS

ANOVA	Analysis of Variance						
A.D.	Anno Domini						
AI	Artificial Intelligence						
ANN	Artificial Neural Network						
BP	Back-Propagation						
B.C.	Before Christ						
BLR	Binary Logistic Regression						
B&W	black and white						
CVA	Canonical Variate Analysis						
CEDAR	Center of Excellence for Document Analysis and Recognition						
CPU	Central processing unit						
COLD	Cloud of Line Distribution						
СТ	Computerized Tomography						
CL	Confidence Level						
CNN	Convolutional Neural Network						
COVID-19	Coronavirus Disease						
DNA	deoxyribonucleic acid						
DFA	Discriminant Function Analysis						
dpi	dots per inch						
e.g.	Example						
L1	first lumbar vertebra						
FDE	forensic document examiner						
FDEs	forensic document examiners						

C4	fourth cervical vertebrae
4IR	Fourth Industrial Revolution
GPA	Generalized Procrustes analysis
GMM	Geometric Morphometric
F	Goodall's F statistic
GPU	Graphic processing unit
GUI	Graphical user interface
HMM	Hidden Markov Model
IT	Information Technology
ΙΟΤ	Internet of Things
Ю	investigating officer
IOs	investigating officers
JPEG	joint photographic experts group
LDA	Linear Discriminant Analysis
MRI	Magnetic Resonance Imaging
MNIST	Modified National Institute of Standards and Technology
MANCOVA	Multivariate Analysis of Covariance
MLP	Multi-Layer Perceptron
MANOVA	Multivariate Analysis of Variance
NB	Naive Bayes
NIST	National Institute of Standards and Technology
NN	Neural Network
NNs	Neural Networks
OS	operating system
OSs	operating systems
OMR	Optical Mark Recognition

PAST	Paleontological Statistics					
PANN	Para-consistent Artificial Neural Network					
PLS	Partial Least Squares					
PC	personal computer					
PDA	personal digital assistant					
px	pixels					
PC	Principal Component					
PCs	Principal Components					
PC1	Principle Components 1					
PC2	Principle Components 2					
PC3	Principle Components 3					
PCA	Principal Component Analysis					
PLC	Programmable Logic Controllers					
SWGDOC	Scientific Working Group for Forensic Document Examination					
SRA	Sekolah Agama Rakyat					
SK	Sekolah Kebangsaan					
SJKC	Sekolah Jenis Kebangsaan Cina					
SJKT	Sekolah Jenis Kebangsaan Tamil					
SRSS	Square root of the sums of squares					
SOPs	Standard Operating Procedures					
SAS	Statiscal Analysis System					
SYSTAT	Statistical Data Analysis and Scientific Visualisation					
SPSS	Statistical Package Social Science					
SOV	subject-object-verb					
SS	sums of squares					
SLLE	Supervised Locally Linear Embedding					

SVM	Support Vector Machine				
i.e.,	That is				
JEPeM	The Human Research Ethics Committee of USM				
3D	three-dimensional				
2D	two-dimensional				
USA	United States of America				
VS	versus				

LIST OF APPENDICES

- Appendix A Infographic from the Department of Statistics Malaysia in 2019
- Appendix B Ethic Approval from JEPEM USM
- Appendix C Recruitment Poster
- Appendix DTwo Different Versions of Questionnaire Forms; in Bahasa
Melayu and English Version
- Appendix E Written Consent Forms in Bahasa Melayu and English
- Appendix F Manual for Tpsutil software
- Appendix G Manual for TpsDig2 software
- Appendix H Manual for MorphoJ software
- Appendix I Manual for Minitab software

PENDEKATAN APLIKASI *GEOMETRIC MORPHOMETRIC* (GMM) UNTUK MENDISKRIMINASIKAN ETNIK PENULIS DENGAN MENGGUNAKAN TULISAN TANGAN ASKARA ANGKA

ABSTRAK

Tulisan tangan adalah unik, di mana tiada dua individu yang berbeza mampu menulis dengan cara yang serupa, meskipun mereka adalah kembar seiras. Setiap individu mempunyai gaya penulisan yang tersendiri, dan variasi tulisan tangan semula jadi mereka juga dipengaruhi oleh pelbagai aspek termasuk warisan budaya atau latar belakang etnik mereka. Namun begitu, mendiskriminasi etnik individu berdasarkan karakter tulisan tangan mereka, apatah lagi aksara angka tulisan tangan, adalah terhad walaupun hakikatnya analisis ini penting dalam membantu pemeriksa dokumen forensic (FDE) dalam menyelesaikan sesuatu jenayah yang berkaitan dengan dokumen yang meragukan. Tujuan umum kajian ini adalah untuk menyiasat kemungkinan untuk mendiskriminasi penulis berdasarkan latar belakang etnik mereka dengan menggunakan tulisan tangan aksara angka penulis dengan menggunakan teknik Geometric Morphometric (GMM). Tulisan tangan aksara angka telah dikumpulkan daripada 390 peserta kajian yang juga datang daripada tiga latar belakang etnik yang berbeza di Malaysia iaitu Melayu, Cina dan India, dan telah didigitalkan dan ditandai dengan menggunakan perisian TpsUtil1.78 dan TpsDig2 sebelum GMM dinilai menggunakan perisian MorphoJ dan Minitab. Dua dan tiga dimensi Analisis Komponen Prinsipal (PCA) menunjukkan terdapat kelompok corak yang boleh dikenalpasti antara penulis Melayu, Cina dan India. Selain itu, terdapat perbezaan yang signifikan dalam angka tulisan tangan 2 hingga 9 dalam kalangan penulis Melayu, Cina dan India apabila set data mereka telah dianalisis

menggunakan Analisis Varian Procrustes (ANOVA) (p<0.0001) dan Analisis Fungsi Diskriminan (DFA) (p<0.0001), yang menunjukkan bahawa aksara angka tulisan tangan tersebut berpotensi untuk digunakan bagi mendiskriminasi penulis berdasarkan etnik mereka.

APPLICATION OF GEOMTERIC MORPHOMETRIC (GMM) APPROACH FOR AUTHOR'S DISCRIMINATION USING THE HANDWRITTEN NUMERAL CHARACTERS

ABSTRACT

Handwriting is unique in which no two individuals can write the same way regardless they are identical or fraternal twins. Everyone has their own unique handwriting style or their natural handwriting variations which are influenced by a variety of factors which include their ethnic background or cultural background. Nonetheless, discriminating individuals' ethnicity based on their handwriting characters, let alone the handwritten numeral characters, is limited even though the handwritten numeral characters analysis can be crucial in assisting forensic document examiner (FDE) in solving crimes related to dubious documents. The general objective of this study was to investigate whether it is possible to discriminate authors based on their ethnic background by utilising their handwritten numeral characters using novel Geometric Morphometric (GMM) technique. The handwritten numeral characters of 0 until 9 were collected from 390 participants from three different ethnic backgrounds in Malaysia which had been digitised and landmarked using TpsUtil1.78 and TpsDig2 software respectively, prior to GMM assessment using MorphoJ and Minitab software. From the results, two-dimensional (2D) and three-dimensional (3D) of Principal Component Analysis (PCA) scatterplots demonstrated identifiable cluster patterns between Malay, Chinese and Indian authors. Besides, there were significant differences in the handwritten numerals 2 through 9 between Malay, Chinese and Indian authors when the datasets tested using Procrustes Analysis of Variance (ANOVA) (p<0.0001) and Discriminant Function

Analysis (DFA) (p<0.0001), which that these handwritten numeral characters have potential to be used to discriminate authors based on their ethnicities.

CHAPTER 1

INTRODUCTION

1.1 Chapter Overview

This chapter describes an overview of handwriting, numeral handwriting, and the Geometric Morphometric (GMM) technique. This chapter aims to introduce these topics and highlight their relevance in the forensic field. Additionally, the problem statement and objectives of the study are also highlighted, emphasising the importance of this research. This serves to inform readers about the general and specific objectives of this study, as well as the challenges that need to be addressed. Overall, this chapter provides an informative and comprehensive introduction to numeral handwriting analysis, making it a valuable resource for anyone interested in this field.

1.2 Study Background

Handwriting is the art or skill of writing that is done by movements of the hand and fingers while gripping a writing instrument, such as pen, marker, or pencil. Besides, handwriting is one of those tool-learned abilities that involves the movement of the whole arm, neuromuscular and complex perceptual-motor habits (Huber and Headrick, 1999). As claimed by Morris (2020), handwriting is not just a handwriting activity but also brain handwriting, in which the brain instructs the arm, hand and fingers as to what should be written or jotted after seeing images or hearing a message. Individuals that are incapable of using their hands, probably could tried to write with their mouths or feet and guided their movements as they attempted to execute the instructions that are sent by the brain (Morris, 2020).

Handwriting is also a part of individual identity and the most advanced achievement that human hands have been capable of, because each individual's handwriting is unique and full with variations as any individuals write to express their feelings and thoughts in a physical way (Huber and Headrick, 1999). In line with Harne *et al.*, (2018), even though handwriting is an art of writing, it is more closely related to a repeating habits of conscious act that required skill to make automatic formation of each letter, character or word before it becomes an involuntary action. In addition, Desai and Kalyan (2013) also stated that technically, handwriting is taught to a person and learned by that person by using a guided or copybook consisting of letter forms, to imitate a letter or combination of letters as it written by someone else. The formations of the letters, characters or symbols within each handwriting system have their own ideal movement patterns as well as spatial relationships between symbols and their directional handwriting (Desai and Kalyan, 2013).

The identification and discrimination of handwriting and numeral handwriting are fundamental in the investigation of forensic questioned document, particularly in cases involving fraud, forgery of signatures or other documents, fabricated checks, disputed wills, forged personal identification numbers, manipulation of birth certificates or any other type of white-collar crimes. It could provide valuable information, and it gives the forensic document examiner (FDE) an overview of how to establish and ascertain the authenticity or authorship as well as how to differentiate between questioned and genuine documents. Furthermore, it also has the potential to be used as crucial evidence in legal proceedings, alongside other types of evidence such as fingerprints, blood, saliva or gunshot residues that have been gathered from the crime scene to identify potential suspects. With the help of this type of evidence, forensic scientist and investigating officer (IO) may be able to make a connection between the victim, the suspect and the crime scene. However, the capability of the forensic document examiners (FDEs) to identify and determine whether forgery has been committed depends heavily on their level of experience, skill and knowledge.

The ability of FDEs to examine, identify and distinguish class and individual characteristics of handwriting and numeral handwriting are based on their judgmental skills, knowledge and experience gained from studying and comparing the copybooks or handwriting of one author with the handwriting of another author (Taylor et al., 2020). This conventional process is frequently used to establish the authenticity of the handwriting and numeral handwriting, as every person has their own distinctive handwriting style that comes in a bundle of variations. This may lead to inaccurate FDE's judgement in court as an expert witness to identify the individual, as well as the possibility for FDEs misidentification of unusual features of alphabet or numeral characters (Taylor et al., 2020). As eloquently stated by Saini, Chauhan and Ganjoo (2020), over the course of the past few years, numerous computational techniques have been developed in the field of handwriting and numeral handwriting analysis for the purpose of facilitating a more expedient investigation process. Because of the handwriting from the sample and questioned documents can be stored in a computer (as offline handwriting) by using a computer's peripheral such as a scanner, this digital advancement has changed and become more specific for the identification, determination and verification of handwriting and numeral handwriting in forged documents (Bensefia and Paquet, 2016; Arbain et al., 2018; Saini et al., 2020).

1.3 Numeral handwriting

Every subfield of mathematics is built on the foundation of numeral systems. According to Lande (2014), numerous numeral character systems have been developed and utilised throughout the history of mathematics, spanning all the periods of civilisation. Numeral handwriting is one of the writing systems that uses a single or a combination of numeral characters to express numbers within a given range by means of writing materials (Gupta and Bag, 2021). To illustrate, a single numeral character is 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9, while a combination numeral character is 10, 11, 12, 13, 14 and so on. Even so, there are many different forms of numeral handwriting systems that use different bases, which eventually converged into modern numbers as presented in Figure 1.1.

Digit	English	Hindi	Bangla	Oriya	Telugu	Gujrati	Punjabi	Arabic
ZERO	0	O	0	0	0	0	\bigcirc	0
ONE	1	9	У	C	c	2	9	2
TWO	2	2	X	9	9	R	2	٢
THREE	3	R	6	$\tilde{\mathbf{\omega}}$	З	3	N	per
FOUR	4	ሪ	8	R	જ	Т	S	ε
FIVE	5	と	(F	R	æ	U	21	ω
SIX	6	Ę	U	5	E	5	S	Ч
SEVEN	7	0	7	D	S	و	2	\vee
EIGHT	8	6	5	7	$\boldsymbol{\sigma}$	6	L	Λ
NINE	9	9	ิง	M	ε	G	S	9

Figure 1.1 The handwritten numeral characters of eight different languages (Bulliet et al., 2019)

Lande (2014) believed that the modern numeral character system was originated from the Hindu-Arabic numeral system, which was developed between the first and fourth centuries. In the beginning, the Indian and Arab mathematicians invented the Hindu-Arabic numeral system and developed the base-ten numeral system earlier (Lande, 2014). Later on, by the seventh century, the Hindu-Arabic numeral character system had spread to the Middle East, Southeast Asia and East Asia including Tanah Melayu (also known as Malaysia) (Bulliet *et al.*, 2019) (Ifrah and Bair, 1987). The Hindu-Arabic numeral system was first introduced in Europe in the early 10th century and this marked the beginning of the modern numeral character system (Bulliet *et al.*, 2019). Since the late 15th century, a wide variety of handwritten and printed numeral characters have been invented, as shown in Figure 1.2 which practically derived from the Roman and Hindu-Arabic base-10 system.



Figure 1.2 Handwritten and printed numerals since the late 15th century (Ifrah and Bair, 1987)

Huber and Headrick (1999) acknowledged that the development of handwriting and numeral handwriting is enigmatic as well. This is because throughout the course of history, countless cultures have emerged, handwriting and numeral handwriting system been passed down through the centuries and constantly changing on a regular basis till this date (Huber and Headrick, 1999). According to Thomas and Rajan (2019), culture, education, orthography and environment of authors are the influential factors in the uniqueness of handwriting and numeral handwriting habits as the individual would be introduced into their culture standard which can lead to distinctive pattern of native linguistic languages. Furthermore, there are numerous instances in which the particular pattern of these movements is cultural, schooling and environment dependent on the habitual aspects of writing which are specific to particular individuals (Huber and Headrick, 1999; Saini and Kapoor, 2018). In addition to this, handwriting and numeral handwriting development process also can be influenced by factors such as one's health, age, handedness, mental state and the surface on which they write.

It is impossible for two individuals, even those who are genetically identical or fraternal twins, to have handwriting and numeral handwriting style that is identical. Furthermore, it was thought that an individual's handwriting, whether be it letter or numeral characters, could be affected by many factors, one of which is their ethnicity. There have been a lot of studies done on how different ethnicities write letters, but there have not been nearly as many done on how the different ethnicities write the numeral characters. The discrimination of individuals based on their ethnicity is still a relatively new idea or concept in the field of forensic questioned document investigation. In order to address this issue, a number of researchers have been attempted to devise and are currently in the process of developing solutions to overcome any circumstances that could lead to the inaccurate discrimination of the handwritten numeral characters between individuals especially when involve with various ethnic groups of authors.

1.4 Geometric Morphometric (GMM)

Morphometric is the study of biological shape variation and its relationship to other variables such as size, length, width, distance ratio and height. Morphometric also employs a multivariate statistical tool to describe the patterns of shape variation within and between groups (Adams *et al.*, 2004). However, a paradigm shift occurred in quantification, which quantified and analysed the data morphological structures (Adams *et al.*, 2004). Adams *et al.*, (2013) proclaimed that the traditional morphological analysis was being supplanted by geometric approaches and the coordinates of anatomical or structure landmarks, while preserving the anatomical or structure information throughout the analysis. In turn, this led to a new revolution in the morphometric research, in which the morphological studies were carried out by calculating linear distances based on the coordinate landmarks and outlines, resulting in a new era of morphometric research (Adams *et al.*, 2004).

Geometric Morphometric (GMM) is a unsupervised machine learning technique for analysing form change and its covariation with other variables by employing two-dimensional (2D) or three-dimensional (3D) points that representing landmarks, curves, outlines or surfaces (Adams *et al.*, 2013). GMM approach was initially used to assess and quantify the variation in the shape of anatomical objects by utilising the Cartesian coordinates of anatomical landmarks after quantitatively accounting quantitatively for the effects of non-shape variation (Adams *et al.*, 2013). In the 1980s, a few fundamental adjustments were made to the GMM in order to focus on the coordinates and their relative geometric positions of landmarks (Adams *et al.*, 2013). As a result of these adjustments, the GMM approach uses univariate and multivariate statistics to quantify shape, coordinate the landmark outlines, curves and surfaces, and would then visualise them as landmark configurations (Adams *et al.*, 2013).

Slice (2007) discussed on how physical anthropology also played an important role in the early development and innovation of the GMM approach. One of the key driving forces behind the advancement of the GMM approach is the complete retention of geometric information from the data collection through analysis and visualisation (Slice, 2007). Since then, the GMM approach has grown into a sophisticated alternative method of shape analysis and now is widely used to quantify anatomical or structure shapes in a wide range of scientific disciplines, including forensic anthropology, entomology, zoology, biology and archaeology (Slice, 2007; Adams *et al.*, 2013; Ibrahim *et al.*, 2019). The GMM approach permits rapid statistical analysis of multivariate data by correlating them to the physical structure, distances, angles, or distance ratios of the original specimens (Slice, 2007). Moreover, GMM data contains 2D or 3D landmark coordinates that correspond to the specimens, and if and only if a specimen differs in shape and size, it will provide distinct landmark coordinates (Webster and Sheets, 2010).

1.5 Problem Statement & Study Rationale

Numerous parties such as legal scholars, civil rights organisations, defence attorneys, scientific community and forensic experts have cast doubt on the credibility of the handwritten numeral characters analysis as a crucial piece of evidence in a court of law, placing it in with the same category as fingerprint, deoxyribonucleic acid (DNA), gunshot residues or blood analysis in terms of its importance. This occurs because the FDE uses a visual examination approach quite frequently to evaluate and analyse the numeral handwriting found on documents that are in question. It is possible there will be a case in court that involves questioned phone number, monetary, wills, account number or secret password that was identified as having been written by suspects. The growing scepticism of handwriting analysis, particularly on the handwritten numeral characters, caused it to be needed in scientific fields, prompting FDE to design, develop and establish a new scientific, statistical and computer-based technique such as Artificial Neural Network (ANN), Multi-Layer Perceptron (MLP), Support Vector Machine (SVM) and Convolutional Neural Network (CNN).

It is widely believed that an individual's handwriting can be influenced by numerous factors, including their ethnicity, culture, gender, handedness, level of education, primary schooling environment and age. It is also possible that these factors can also influence the handwriting style used for numeral character expressions. According to the findings of a study conducted by Cheng *et al.* (2005), the FDE is able to identify and determine the nationality or ethnic origins of authors based on several classes and individual characteristics when the handwritten numeral characters specimens have been subjected into a comprehensive examination. However, even though FDE analyses the authors' the handwritten numeral characters based on the class and individual characteristics of the handwritten numeral characters, FDE is still unable to achieve a high percentage of Confidence Level (CL) of findings related to the authors' country of origin or ethnic background. This is the case even with the help of ANN, CNN, SVM and Cloud of Line Distribution (COLD) that have been made in the field of forensic sciences in order to analyse the handwritten numeral characters between different authors. This study proposes an exploratory study, which also known as an observational study, that focuses on the discrimination of the handwritten numeral characters between three main ethnic groups in Malaysia; Malay, Chinese and Indian by using the novel Geometric Morphometric (GMM) technique. GMM approach offers assessment of the shape variation based on the Cartesian geometric coordinates by utilising the specimens' configuration landmarks of the handwritten numeral characters. PCA scatterplot, Procrustes ANOVA and DFA results were be used prior to demonstrating and visualising the pattern of shape variations of the handwritten numeral characters and comparing the covariation between the specimens' geometry morphology and other continuous variable such as ethnic groups.

As far as this study is concerned, research that focuses on identifying, determining and discriminating the ethnicity of individuals based on their handwriting characters, let alone the handwritten numeral characters, is limited elsewhere including in the Malaysia setting. Malaysia is a multi-racial country where three main ethnic groups namely Malay, Chinese and Indian made up 90 percentage (%) of its population. It is a widely held belief that individuals' handwriting is influenced by many factors, one of which is their ethnic background. This is due to the fact that individuals from the same ethnic group have the potential to write the numeral characters in a similar ways and develop certain level of similarity in their numeral handwriting attribute as the range of variation in handwriting within the same ethnic group is minimal in comparison to the range of variation in handwriting amongst individuals from different ethnic groups (Deepani and Kapoor, 2018). In a case of document fraud involving multiple persons of different ethnicities, FDE

could implements this novel GMM approach to identify and discriminate individuals based on their handwritten numeral characters, which could be used as evidence.

1.6 Research Question(s)

This study aimed to offer a scientific and computer-based analysis for the handwritten numeral characters of Malaysia's three main ethnic groups. This study implemented the GMM approach to discriminate the ethnicity of authors based on their numeral handwriting. The main research question to be pondered about is whether it is possible to discriminate the ethnicity of authors based on their handwritten numeral characters. In order to answer this question, each of the handwritten numeral character between Malay, Chinese and Indian authors was analysed and compared in an effort to identify any distinguishing class characteristics (mainly about the shape formation) prior to landmarking process.

In this study, statistical techniques such as PCA, Procrustes ANOVA and DFA were also employed in order to find any measurable discrimination between different ethnicities of authors based on their handwritten numeral characters. However, which appropriate statistical techniques that actually should be performed in this study to analyse and visualise the findings? Are there any possible circumstances that could be arised if the chosen designated statistical techniques are used to evaluate and assess the findings? Last but not least, how precise are this designated statistical techniques and computer-based approach that can be utilised scientifically to differentiate between Malay, Chinese and Indian authors?

1.7 Significance of the Study

This study provides the opportunity to contribute to the information regarding the manner in which the handwritten numeral characters could be used to discriminate the authors between the three main ethnic groups in Malaysia, namely Malay, Chinese and Indian. Discrimination between individuals based on their handwritten numeral characters limited globally, especially in multi-ethnic countries, including Malaysia, despite the fact that 90% of the population is comprises Malay, Chinese and Indian individuals. This study has the potential to be a reliable indicator for discriminating the ethnicity of authors by examining and analysing their shape variation of handwritten numeral characters. GMM approach is simple in practice which can be used scientifically in the courtroom to assist an IO and FDE if they come across a suspicious document containing signs of forgery or alterations.

1.8 Objectives

1.8.1 General

The general objective of this study is to use the specimens of the handwritten numeral characters collected from the three main ethnic groups in Malaysia; Malay, Chinese and Indian, in order to discriminate the ethnicity of the authors by utilising the GMM approach.

1.8.2 Specific Objectives

The specific objectives of this study are as follows:

- To utilise the viability of the novel Geometric Morphometric (GMM) approach for discriminating the ethnicity of the authors based on their handwritten numerals characters.
- To establish identifiable landmark configurations which representing the geometry morphology of the handwritten numeral characters, onto the Procrustes coordinate system for digitisation process of specimens.

- 3. To generate and analyse the shape variations that can be observed in each handwritten numeral character with the continuous variable; ethnic groups, between Malay, Chinese and Indian authors in Malaysia setting using the GMM approach.
- To evaluate the relationship of the handwritten numeral character with the ethnicity factor by utilising statistical techniques; Principal Component Analysis (PCA), Procrustes Analysis of Variance (ANOVA) and Discriminant Function Analysis (DFA).

CHAPTER 2

LITERATURE REVIEW

2.1 Chapter Overview

This chapter provides an overview of several literature reviews on the uniqueness about handwriting between individuals, along with the logical approach to the handwriting as evidence, are emphasized and critically evaluating the quality and validity of the studies reviewed. Furthermore, this chapter reviews some factors that could influenced handwriting style and the advances made in the forensic discipline regarding the identification and discrimination of handwritten numeral characters in order to identify the gaps or areas that require further investigation. Studies on the GMM approach and statistical analysis are also included in this context for this study's theoretical framework and methodology.

2.2 The Uniqueness of Handwriting

Handwriting is an intricate neuromuscular interplay that originates in the central nervous system to perform the writing task. Authors' hands, arms, feet and mouths are all coordinated in unison to carry out information from the brain, which being transmitted through nerves via the neurological system (Amend and Ruiz, 1980; Morris, 2020). In the simplest definition, handwriting is a physical act motivated by the mind's creative process and intended to produce written word forms through the pen or pencil's movement (Amend and Ruiz, 1980; Tarannum *et al.*, 2015). The brain's instructions comprise information about the appearance of the alphabets, letter combinations, numeral characters and other symbols as well as how they should be written (Morris, 2020). A person's ability to manipulate and control a writing instrument can dictate the style of their handwriting (Morris, 2020).

Handwriting also can be regarded as a visual medium, owing to the fact that it is complicated due to the articulation tactics used for alphanumeric text and visual imagery developed through practice and skill movement (Neef *et al.*, 2006). During childhood, any individuals learn how to write by copying the copybook at primary or vernacular school before adapting it to their own 'individual habits' until they attained the graphic maturity stage of development. Due to the fact that humans are not functioning like a robot, their own handwriting style does not always look exactly the same all the times due to some natural variations and cause to develop their own individual characteristics in the handwriting which distinguishable with another individuals' handwriting (Stewart, 2017). Besides, the writing instrument's coordinated movement and pressure could lead across the page, producing an unchangeable and personal imprint of handwriting style (Neef *et al.*, 2006). Since handwriting is a continuous and flowing task, the writing instrument movement may occasionally be uninterrupted, even if not recorded as an inked line (Harralson and Miller, 2018).

The individual's style and unique pattern of the handwriting movement occurring at the appropriate time, place and purpose, which defines that every individual has its own personalised touch (Neef *et al.*, 2006; Mohamed *et al.*, 2010). As handwriting habits become more automatic with passing of time, the handwriting process becomes less susceptible to conscious control and a person would not write like how he or she did during primary school (Neef *et al.*, 2006). This is because the authors tend to focus on what vital message that they are trying to write, rather than to focus on the handwriting process. Besides, handwriting movements also involve several subconscious habits and makes the handwriting style is distinct and peculiar due to their recurring patterns (Neef *et al.*, 2006). Handwriting style and individual

characteristics can be used to distinguish our handwriting with someone else's, in order to determine the authorship based on the degree of similarities, accuracy and certainty when the specimens compared.

Even with the development and emergence of sophisticated writing machines such as a typewriter machine, smartphone, tablets or computer with a touchscreen to write and documented their documents, handwriting has remained as an important part of individual's lives (Neef *et al.*, 2006; Bensefia and Paquet, 2016; Khushboo *et al.*, 2020). Despite the fact that the digital age has been with us for decades, some individuals still prefer the manual handwriting to write their personal and factual information (Khushboo *et al.*, 2020) rather than to write and save it by online due to fear and anxiety about the cyber thieves. This contributes to the growing number of the handwritten documents that surround us on a daily basis (Bensefia and Paquet, 2016; Angel and Kelly, 2021). Aside from that, as our modern writing machines evolved, so did our handwriting's practical functions and social meanings as well as its cultural aesthetic values (Neef *et al.*, 2006).

As a result, handwriting's use and meaning probably have shifted over time in response to advancement of technological, social and cultural transformations (Neef *et al.*, 2006). Hence, handwriting is very unlikely to disappear in the future, as long as its technologies are closely linked to specific cultural activities and official forms that are still used today (Neef *et al.*, 2006). However, forensic handwriting analysis always subject to the controversies since their subjective analysis is mainly depend on the visual inspection of the handwriting (Bensefia and Paquet, 2016). Due to that, FDE tried very hard to develop specific tools with advancement technologies using artificial intelligence (AI) in order to analyse a huge quantity of documents based on the analysis of specific need (Bensefia and Paquet, 2016). As a consequence, there

are many studies where the AI and machine learning technique were applied into the discipline of the questioned document examination against unlawful activities (De Alcaraz-Fossoul and Roberts, 2017).

In the forensic science, handwriting analysis is one of forensic analyses which unique and very subjective. The fact that there are no two handwritings are exactly alike in appearance, but unique to each individual, is what qualifies the handwriting as an crucial evidence for the court use (Huber and Headrick, 1999). Even an American statistician concluded that the probability of two handwritings being identical is one chance in 68 trillion or less (Amend and Ruiz, 1980). In fact, handwriting also can be used for identification and verification in the same way that fingerprints and DNA used by the forensic investigators for analysis and absolute identification (Stewart, 2017). Stewart (2017) asserted that it is impossible for two individuals (including identical or fraternal twins) to have identical fingerprints and handwriting, but some studies suggest that some similarities probably could be existed but there is no research or study to validate scientifically on this statement.

The researchers also should consider that the handwritten numeral character would be a great type of evidence for the forensic investigation as FDE could be encountered it on the suspicious bill, bank check, mail, official form, password or coding number secret code. Identification and discrimination of the handwritten numeral characters are helpful to reduce the case of fraud in commercial transactions or aid to solve an investigation involving with numbers, yet the idea of this study to perform discrimination of the handwritten numeral characters between different ethnicities are due to the limited scientific and computer-based study. Besides, some studies on the handwritten numeral characters mostly were conducted based on the researcher's subjective knowledge, skill, reference database and using some conventional instruments such as stereomicroscope, video spectral comparator or conventional magnifying class. Bojja *et al.* (2019) and Sharma *et al.* (2020) stated that the aims to develop handwriting character recognition and discrimination software, were to get a higher accuracy rate of character analysis, reduce time consuming and space, decrease manual paperwork, offer the higher degree of match between a questioned and known document prior to making it optimal compared to the manual methods of the questioned document analysis. At the same time, the authors also mentioned that, in order to avoid handwritten document become distorted by time and can no longer useful for court or investigation use, its best to convert it into digital image format as offline handwriting which could reduce time consuming of retrieval process and make handling process become more easier and reliable (Bojja *et al.*, 2019; Sharma *et al.*, 2020).

2.2.1 Handwriting as an Evidence

Almost every type of criminal or civil cases will entail the questioned documents, particularly those involving the white-collar criminals. Handwriting, offline or digital signatures, computer fonts, printing and security features on documents are the most frequently requested for the questioned document examinations (Angel and Kelly, 2021). When there is a suspect involved in a crime and the collected evidences are including a handwritten document, IOs may contact the FDE to analyse and examine if there is a match between the evidence and the known specimen (Angel and Kelly, 2021). A trained FDE will assess the handwriting evidence and the known specimen prior to determine the potential authorship, identify the authenticity or non-authenticity handwriting and disclose if there is any alterations, additions, erasures, or deletions made on the handwriting evidence (Angel 2021). and Kelly, In circumstances. the only one rare

piece of handwritten evidence may be sufficient to bring the suspect to the trial and conviction (Angel and Kelly, 2021).

Occasionally, FDE examines, analyses and compares the questioned documents that containing handwriting characteristic which discovered at the crime scene, with the known reference specimens for a number of reasons. For instance, FDE needs to determine whether the handwriting from questioned document, was or was not from the suspect's known writing specimens prior to offering their testimony in court as an expert (Amend and Ruiz, 1980). The handwriting expert's testimony is admissible in the court of law for the purpose of establishing identification or authenticity of the particular documents (Amend and Ruiz, 1980). Generally, FDE works in cooperation with law enforcement agencies, attorneys or anyone who is interested in identifying 'who wrote that, when, where and under what circumstances' (Amend and Ruiz, 1980).

The analysis of handwriting evidence relies upon the notions that there is no two individuals could write identically, even no one could writes exactly in the same manner twice and the chance of two authors having the same profile is extremely low (Mohamed *et al.*, 2010; Saunders *et al.*, 2011; Tarannum *et al.*, 2015). The identification of class and individual characteristics by the FDE at the handwriting examination stage, is a very crucial step in determining a person's identity (Stewart, 2017) and even the study by Saunders *et al.* (2011) was considering that the degree of individuality of authors on 98 authors before randomly selecting two different individual with same writing profile using Random Match Probability (RMP), was higher in value. However, it is possible for an individual to purposely change his or her handwriting style in order to achieve a different appearance for a number of different reasons (Tarannum *et al.*, 2015). If so, FDEs should be able to articulate a comprehensive set of characteristics as that might fairly to be expected when they dealing with disguised, simulated or traced writing, as well as genuine handwriting in various scenarios and propositions (Angel and Kelly, 2021). Aside from that, it is widely agreed by FDE that the level of education, schooling environment, cultural, gender, ethnic background, physical health, the writing instruments and type of writing surface have an impact on the written result (Tarannum *et al.*, 2015; Johnson *et al.*, 2017; Nag *et al.*, 2018; Sahani, 2021).

Forensic handwriting analysis is a complicated, time-consuming and methodical process that requires a thorough understanding on how people write the sentences using the combination of alphabets and numerals (Angel and Kelly, 2021). Because of the distinctiveness of their handwriting, which can be affected by their physiological processes and motor fine skills, a person's handwriting can reveal insights about his or her identity (Angel and Kelly, 2021). Even in the rare instance, where an individual has demonstrates significantly different handwriting style, it is possible to conclude that the handwriting styles share critical identifying class or individual features after applying forensic methods when examining and comparing handwriting characteristics in one-to-one manner (Tarannum et al., 2015; Deepani and Kapoor, 2018; Tyagi, 2018; Gannetion et al., 2022). Any handwriting evidence is thoroughly examined for signs of manipulation, alterations, and inconsistencies in the handwriting style prior to writing an objective report on it (Angel and Kelly, 2021). Because of the FDE is the ones who advocate for the handwriting evidence, the FDE must realise that their examination and comparison techniques need not to change when they applying the logical approach during examination (Taylor et al., 2020; Angel and Kelly, 2021).

Logical approach is a framework for the proper evaluation of findings derived from the questioned document examination and act as a guidelines for reporting the results either written or verbal form (Ostrum, 2019). This logical approach is important because during the examination of handwriting evidence, it is possible for the FDE to commit a mistake (Saunders *et al.*, 2011; Thomas and Rajan, 2019). Occasionally, FDEs should considered the presence of the natural variations in handwriting as a disguised or forged handwriting, which elicit them to make a report upon this error (Thomas and Rajan, 2019). If there are two writing specimens from different individuals that are declared to be 'match', then the false match error has occurred, meanwhile when the false 'no-match error has occurred if the two writing specimens from the same individual are declared to be 'no match' (Saunders *et al.*, 2011; Thomas and Rajan, 2019). On the other hand, the FDE possibly to arrive the 'no conclusion' decision if the writing specimens are insufficient to comprehensively determine whether or not the two writings specimens were written by the same individual because no result is also a result (Saunders *et al.*, 2011; Ostrum, 2019).

The objective opinion of the FDE not only must be based on logic, robustness and reliable knowledge, but it must also be transparent (Ostrum, 2019; Angel and Kelly, 2021). This point relates to the issues of disclosure, as the FDE must present all the relevant information that forms the basis of their opinion in its entirety (Angel and Kelly, 2021). During examination process on the handwriting evidence, the FDEs must be focused on the given the propositions and pertinent background information, rather than expressing any belief personally about the likelihood of the propositions themselves (Angel and Kelly, 2021). By expanding testimony to include statements about the propositions, requires the FDEs to go beyond the scope of their expertise, which is contrary to the logical approach's philosophy (Ostrum, 2019; Angel and Kelly, 2021).

2.3 Factors Influenced Handwriting

Research on handwriting is surprisingly promising in this discipline where further developments will be achieved with the advancement of technology in order to remain relevant with the Fourth Industrial Revolution (4IR) (Topaloglu and Ekmekci, 2017). An individual's handwriting is a movement of habitual patterns that are blend with an individual's personality and the things that they have been taught since childhood time and inherit the cultural bias (Mohamed *et al.*, 2010; Saunders *et al.*, 2011; Sahani, 2021). There are so many factors that can influence the development and style of the handwritten numeral characters of an individual. According to the Huber and Headrick (1999), the key factors that can cause variations in the handwriting and numeral handwriting are the writing position, writing surface and writing instrument. Conversely, education level, gender, religion, socio-economic status, age, culture, ethnicity, physical and spiritual changes all have an affect the handwriting performance and make a noticeable appearance of the hand and numeral handwriting (Tarannum *et al.*, 2015; Topaloglu and Ekmekci, 2017; Saini and Kaur, 2019; Gannetion *et al.*, 2022).

It is true that differing positions or stances can have an impact on how someone writes. It was discovered that writing when bending, leaning, sitting in a confined place, walking or holding some weights could result disorganised handwriting when been compared to write in standing position (Khushboo *et al.*, 2020). Sharma and Azeen (2015) and Abbasi *et al.* (2009) even undertook a study in which specimens were collected under the unconventional circumstances, such as

22

writing while travelling in automobiles or walking, in order to prove that writing position could influenced the hand and numeral writing.

Khushboo *et al.* (2020) explored on how different types of writing instruments could make observable differences in the handwriting across individual due to their personal preference for utilising it. This statement was supported by the study of Tarannum *et al.* (2015) on the handwriting from 50 people using different unconventional writing instruments such as lipstick, charcoal, eyeliner or crayon, which then revealed that the appearance of the handwriting could varies in terms of size, length and positions of letters, characters or words, relative ratio, breadth of letters, and spacing between letters. This study also mentioned that it could lead to error findings if the different individuals write numeral characters differently when they used unconventional writing instruments (lipstick, Kajal or Coal) under unconventional stances or positions (walking or standing) as compared to when they write using conventional writing instrument (pen, pencil or marker) under conventional stances or positions (sitting or lying).

Khushboo *et al.* (2020) also disclosed that the questioned document that contains suspicious handwriting which could be found on unusual writing surfaces, is a difficult task for the FDEs to examine and analyse. This is because in some case, the pattern of the handwriting from the same individual might appear differ and vary depending on the types of writing surfaces use such as on the roof car, the rocky wall, the kin surface or on a mirror (Khushboo *et al.*, 2020). However, according to the study by Tarannum *et al.* (2015), it was revealed that there is no significant difference in the handwriting of any individual owing to change in the writing surfaces on the basis of ANOVA analysis.

On the other hand, because of the hand and numeral writing is a reflection of the author's state of mind at the time it is used to convey the instruction of brain (Amend and Ruiz, 1980), the pattern of hand and numeral writing which can distinguishes one person's script from another, could be influenced by the author's perceptual abilities physiology process, intellectual development, mental and state of health of the authors (Tarannum *et al.*, 2015; Topaloglu and Ekmekci, 2017). This is because the spiritual and physical disorders such as anxiousness, tension or drugs intoxication tend to create trembling and cause pen slippage. This could bring to distortions in spacing, relative size of strokes, letter formations and directions.

Generally, hand and numeral writing are an important fine motor skill of learning during the school age, with more than 50% of students engaged in handwriting tasks at the school (Shaturaev, 2019; Gannetion et al., 2022). While learning to write, any individuals are introduced to cultural handwriting norms such as neatness and regularity, as well as the cultural biases like write from right-to-left or left-to-right and counter-clockwise rotations (Huber and Headrick, 1999; Cheng et al., 2005; Kapoor and Saini, 2017). The grasp, posture and the hand used to write, are some of the other constraints that may differ as per the standards of a culture to another different culture, which resulting in varied effects in handwriting style (Huber and Headrick, 1999; Kapoor and Saini, 2017). On top of that, Mohamed et al. (2010) and Stewart (2017) proclaimed that handwriting is the combination of class and individual characteristics. Roughly, class characteristics usually happened resulting from writing system studied, education, ethnic background and family associations, while individual characteristics are uncommon to be happened in handwriting as it would be done either consciously or unconsciously act in only one person's writing (Mohamed et al., 2010; Deepani and Kapoor, 2018).